

University of Groningen

Leerlingen – klassen – scholen

Veenstra, René

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

1999

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Veenstra, R. (1999). *Leerlingen – klassen – scholen: Prestaties en vorderingen van leerlingen in het voortgezet onderwijs*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

1 | Inleiding

1.1 Scholen op rapport

Kunnen de prestaties van scholen zo maar worden vergeleken? Wanneer geeft een ranglijst van scholen een eerlijk en goed beeld? Zijn er tussen scholen grote verschillen in prestaties? In hoeverre zijn verschillen tussen scholen toe te schrijven aan de school of aan docenten? Of komen die verschillen vooral door kenmerken van leerlingen en gezinnen? Welke leerlingen gaan vooruit in het voortgezet onderwijs en welke leerlingen raken achterop? In welke mate zijn deze verschillen het gevolg van ongelijke onderwijskansen? Zijn er kenmerken waardoor deze ongelijke onderwijskansen kunnen worden verminderd? Op dergelijke vragen wordt in dit boek een antwoord gegeven.

De belangstelling voor verschillen in prestaties werd vrij recent aangewakkerd door het landelijke dagblad *Trouw*. Eind 1997 maakte het dagblad de prestaties van alle scholen voor voortgezet onderwijs bekend. In paragraaf 1.1 van deze inleiding wordt de inhoud van dit Trouw-onderzoek uiteengezet. Door met name onderwijskundigen, sociologen en psychologen is veelvuldig onderzoek gedaan naar verschillen in prestaties. In paragraaf 1.2 geef ik een overzicht van dat onderzoek. In die paragraaf presenteer ik ook een kader, de zogeheten *structuur van concentrische cirkels*, waarmee ik verderop in dit boek verschillen in prestaties probeer te verklaren. In paragraaf 1.3 spits ik dit onderzoek toe en geef ik de afbakening, de probleemstelling, de relevantie en de opzet weer.

Efficiëntie, toegankelijkheid en sturing

Op de kosten en de efficiëntie van het onderwijs legt het ministerie van Onderwijs, Cultuur en Wetenschappen sinds de jaren tachtig een sterke nadruk. Een van de centrale begrippen in dat onderwijsbeleid is bijvoorbeeld *schaalvergroting*. In het voortgezet onderwijs was de invloed van die schaalvergroting duidelijk te merken. Het aantal scholen ging van 2.472 in 1980 naar 803 in 1995 (CBS, 1997). Ook het beleid van de overheid voor de *loopbanen* van leerlingen is de laatste jaren steeds zakelijker geworden. Enige voorbeelden hiervan zijn het verhogen van de eigen bijdrage van deelnemers aan het voortgezet, middelbaar en hoger onderwijs; het verbieden dat een leerling in het voortgezet onderwijs een schooljaar twee keer overdoet; het invoeren van profielen in de bovenbouw van havo en vwo

die elk tot slechts een deel van het hoger onderwijs toegang geven; het terugdringen van de doorstroom van het hoger beroepsonderwijs naar de universiteit; het beperken van de studieduur in het hoger onderwijs tot vier jaar. Dergelijke maatregelen duiden erop dat het ministerie de schoolloopbaan van leerlingen efficiënter wil laten verlopen.

Het ministerie wil ook dat ouders en overheid zicht hebben op de kwaliteit van een school. Scholen moeten zich publiekelijk gaan verantwoorden voor hun prestaties. Het ministerie beoogt op die manier ouders beter voor te lichten, de resultaten van scholen vergelijkbaar te maken en scholen te prikkelen tot betere prestaties. Het belang van de overheid bij verantwoording door scholen is de laatste jaren toegenomen, doordat scholen meer autonomie hebben gekregen. Nu de overheid minder vastlegt wat scholen moeten doen, wil de overheid wel meer inzicht in 'wat scholen er toe doen'.

In het hoger onderwijs heeft deze verantwoordingsplicht geleid tot de instelling van visitatiecommissies. Ook basisscholen hebben te maken met een soort visitaties, het zogeheten Integraal Schooltoezicht. Op basis van een bezoek van drie dagen beoordeelt de Inspectie van het Onderwijs de didactiek, het lesmateriaal, de netheid, de sfeer, de samenwerking tussen docenten, de leerlingenzorg, de effectieve leertijd en het pedagogisch klimaat van alle basisscholen. Verder heeft het ministerie het plan om de doorstroomcijfers van basisscholen bekend te gaan maken: welk percentage van de leerlingen gaat naar welk type van voortgezet onderwijs? In het voortgezet onderwijs heeft het ministerie zogeheten kwaliteitskaarten ontwikkeld. De ontwikkeling van deze kwaliteitskaarten kwam in een stroomversnelling toen het dagblad *Trouw* prestaties van alle scholen voor voortgezet onderwijs in Nederland openbaar maakte (Agerbeek, Hageman, Kreulen & Lakmaker, 1997).ⁱ

Deze aandacht voor de kwaliteit van onderwijs valt ook internationaal waar te nemen. In *Education at a glance* vergelijkt de Organisatie voor Economische Samenwerking en Ontwikkeling (OECD, 1992 en later) de schoolprestaties in verscheidene westerse landen.

Eerste Trouw-onderzoek naar schoolprestaties

In het najaar van 1997 maakte *Trouw* in een bijlage de schoolprestaties van alle scholen voor voortgezet onderwijs in Nederland openbaar (Agerbeek e.a., 1997). Geordend naar provincie en schooltype publiceerde het dagblad voor elke school het percentage 'onvertraagd geslaagden' (een schatting van het percentage leerlingen dat het eindexamen haalt zonder in het voorlaatste leerjaar te zijn blijven zitten), het percentage uitvallers en zittenblijvers op de school in de jaren daarvoor en het gemiddelde van de examencijfers voor Nederlands, Engels en Wiskunde A.

Deze gegevens werden voor vbo, mavo, havo en vwo weergegeven. Met deze gegevens berekende de onderwijssocioloog Dronkers ten slotte een rapportcijfer. Behalve de voorgaande prestatiematen werd in die berekening ook het percentage allochtonen betrokken. Door dit percentage mee te wegen probeerde Trouw te voorkomen dat scholen met een 'zwarte' leerlingenpopulatie als vanzelf tot de slechtste worden gerekend.

Aanvankelijk weigerde het ministerie Trouw de gegevens over schoolprestaties te verstrekken. De weigering was gebaseerd op twee gronden. Het doorgeven van het materiaal aan derden zou de vertrouwensrelatie van de inspectie met de scholen aantasten. Ten tweede zouden de scholen schade ondervinden van publicatie van de gegevens, omdat slagingspercentages geen rekening houden met de achtergrond van leerlingen. De redactie van het dagblad ging vervolgens naar de rechter en deed een beroep op de Wet Openbaarheid van Bestuur. De rechter stelde de krant in het gelijk. De vertrouwensrelatie van de inspectie met de scholen kon volgens de rechter niet in gevaar zijn, omdat scholen wettelijk verplicht zijn de inspectie gegevens te leveren. Verder vond de rechter het publieke belang van het vrijgeven van cijfers zwaarder wegen dan de bedenkingen die het ministerie uitte.

In Engeland en Frankrijk zijn schoolprestaties openbaar en bestaan er al langer 'hitlijsten' van scholen.ⁱⁱ Per onderwijsregio publiceert de Engelse overheid jaarlijks de resultaten van scholen voor voortgezet onderwijs. In die publicatie staan twee maten centraal, het percentage leerlingen dat een minimum aantal certificaten voor voortgezet onderwijs (*general certificate of secondary education*) heeft behaald en de score van leerlingen op de bijbehorende toetsen. Verscheidene dagbladen zetten deze cijfers om in *league tables*, waarin de prestaties van scholen op een rij worden gezet. Het afgelopen jaar verscheen ook een overzicht van scholen waar de prestaties de laatste jaren het meest zijn verbeterd. Behalve de resultaten van het voortgezet onderwijs geeft de Engelse overheid sinds vorig jaar ook die van basisscholen weer.

De kloof tussen goede en slechte scholen kan door dergelijke hitlijsten groter worden (zie ook Bosker, Lam, Luyten, Steen & De Vos, 1998, hoofdstuk 2). Zo zou kunnen worden verondersteld dat ouders hun kinderen het liefst aanmelden bij scholen met een hoog percentage geslaagden. Vervolgens zouden deze scholen hun reputatie met gemak in stand kunnen houden door potentieel minder goede leerlingen te weigeren of tussentijds te laten vertrekken. Vandaar dat Trouw het nodig vond om niet alleen de slagingspercentages maar ook andere cijfers in de krant te zetten.

Op grond van het Trouw-onderzoek mocht de Stedelijke Scholengemeenschap Roermond zich in 1997 de beste school van Nederland noemen. De school kreeg

een negen, een negen en een acht voor respectievelijk de vwo-, havo- en mavo-afdeling. Op bijvoorbeeld het vwo slaagde 97 procent van de leerlingen zonder in de vijfde klas te zijn blijven zitten. Het gemiddelde eindexamencijfer was een 7,0 en het percentage uitvallers en zittenblijvers 1 procent. Deze fraaie cijfers behaalde de school uit Roermond met 11 procent allochtone leerlingen. Verscheidene scholen kregen daarentegen een zware onvoldoende als rapportcijfer van Trouw. Op basis van deze cijfers kregen scholen in grote steden als Rotterdam en Utrecht, maar ook in kleine plaatsen als Lichtenvoorde en Sint Michielsgestel het predikaat 'slecht' opgeprikt.

De weergave van deze cijfers omlijstte Trouw met artikelen waarvan de koppen niets aan duidelijkheid overlieten, zoals 'Montessorischolen scoren slecht' of 'Mythen doorgeprikt' (Agerbeek e.a., 1997, p. 1): "Over de kwaliteit van scholen doen vele indianenverhalen de ronde: veel allochtonen op school verminderen de prestaties, categorale gymnasia zijn beter dan (...) scholengemeenschappen en openbare scholen zijn slecht. In wetenschappelijk onderzoek zijn die veronderstellingen vaak getoetst op basis van steekproeven, met soms wisselende conclusies. Nu alle scholen in de berekeningen kunnen worden meegenomen, blijken heel wat mythen doorgeprikt te kunnen worden."

Reacties op het Trouw-onderzoek

Op het Trouw-onderzoek kwamen positieve en negatieve reacties. In de eerste weken na de publicatie stonden er in vele dag- en weekbladen commentaren en ingezonden brieven. Ook op radio en televisie was er volop aandacht voor het onderzoek. De Trouw-redactie kreeg lof, omdat de resultaten van verschillende scholen nu eindelijk openbaar werden. Ook constateerde menigeen met vreugde dat de redactie niet alleen de slagingspercentages maar ook andere uitkomsten van scholen in de beoordeling betrok. Anderen brachten naar voren dat scholen nu beter konden worden gecontroleerd en juichten het toe dat het afgelopen was met de beschermde status van docenten.

Een groot deel van de reacties was echter negatief over het onderzoek. Zo werd Trouw tendentieuze berichtgeving verweten, zouden de gebruikte gegevens nogal wat onjuistheden bevatten en ontstond een discussie over schoolprestaties als een te beperkte interpretatie van wat goede scholen hun leerlingen bieden. Andere commentaren richtten zich op het rapportcijfer dat aan scholen was toegekend en de geringe inzichtelijkheid daarvan, en de vraag of voldoende rekening werd gehouden met de soms aanzienlijke verschillen die tussen leerlingenpopulaties bestaan (zie voor een bespreking van de kritiek: Veenstra, Dijkstra, Peschar & Snijders, 1998).

Kwaliteitskaart

De onderwijsinspectie heeft in het najaar van 1998 voor het eerst een kwaliteitskaart voor scholen voor voortgezet onderwijs uitgebracht.ⁱⁱⁱ Per school staan op deze kaart voor alle schooltypen van elke locatie kerngegevens vermeld. Het gemiddelde examencijfer is een van die gegevens. Dit cijfer wordt op zes manieren weergegeven, namelijk voor alle vakken en afzonderlijk voor Nederlands, talen (Engels, Frans en Duits), aardrijkskunde en geschiedenis, economische vakken en bèta-vakken (biologie, natuurkunde, scheikunde, wiskunde). Voor de eerste drie leerjaren wordt aangegeven hoeveel procent van de leerlingen niet blijft zitten. Voor de hogere leerjaren, dat wil zeggen vanaf de derde klas, wordt berekend welk percentage leerlingen zonder vertraging een diploma haalt. Ook vermeldt de kaart hoe groot de klassen zijn. In de toekomst wil het ministerie ook vermelden wat het ziekteverzuim van docenten is, welke maatregelen worden genomen bij lesuitval en tussenuren, hoeveel leerlingen de school verlaten zonder diploma, en hoe leerlingen presteren in het vervolgonderwijs.

De Onderwijsraad (1998), het adviesorgaan voor de regering op het terrein van onderwijs, heeft gesteld dat de overheid zich bij het berekenen van kwaliteit moet baseren op controleerbare, objectieve gegevens. Om dat te bereiken zouden alle leerlingen een aanvangstoets moeten maken. De resultaten op die toets zouden dan kunnen worden vergeleken met de prestaties aan het einde van het voortgezet onderwijs. Zo lang een dergelijke begintoets er nog niet is, moet de overheid volgens de Onderwijsraad niet een toevlucht nemen tot vage, onduidelijke gegevens, bijvoorbeeld het percentage allochtone leerlingen of de urbanisatiegraad. Dit adviesorgaan vindt dus dat de kwaliteitskaart van het ministerie aan hogere standaarden moet voldoen dan de beoordeling uit Trouw. Verder vindt de Onderwijsraad dat het rendement van scholen op de kwaliteitskaart centraal moet staan en dat alleen zaken moeten worden opgenomen die wettelijk zijn vastgelegd, de zogeheten deugdelijkheideisen.

Netto en bruto prestaties

Het Trouw-onderzoek naar schoolprestaties en de ontwikkeling van kwaliteitskaarten past in een trend om zichtbaar te maken 'wat scholen ertoe doen'. Wanneer de prestaties van scholen zichtbaar worden gemaakt, zal er rekening moeten worden gehouden met de *omstandigheden* waarin scholen verkeren. Een vwo-school met leerlingen waarvan de ouders veelal hoger onderwijs hebben gevolgd heeft een andere leerlingenpopulatie dan een school met leerlingen waarvan de ouders voornamelijk lager beroepsonderwijs of mavo hebben gedaan. Een vbo-school waarvan de meeste leerlingen in potentie het C-niveau of hoger aankunnen verschilt van een vbo-school met leerlingen die alleen geschikt zijn voor het A- of

B-niveau. Om scholen eerlijk met elkaar te kunnen vergelijken, moet er dus rekening worden gehouden met de samenstelling van de leerlingenpopulatie van scholen.

Bij het vaststellen van de prestaties van scholen is het gebruikelijk om een onderscheid te maken tussen *netto* en *bruto* prestaties. De bruto prestatie van een school is het gemiddelde niveau dat leerlingen op een bepaalde maat hebben, bijvoorbeeld het niveau van leerlingen op een toets wiskunde. De netto prestatie van een school is het niveau dat leerlingen gemiddeld hebben, nadat er rekening is gehouden met de samenstelling van de leerlingenpopulatie van de school (Bosker, 1990). In bruto prestaties kunnen de verschillen tussen scholen groot zijn. Vooral in landen, zoals Nederland, waar de leerlingenpopulatie van scholen sterk verschilt, zijn de verschillen aanzienlijk. Wanneer de prestaties van scholen echter op een eerlijke manier met elkaar worden vergeleken en er dus netto prestaties worden berekend, dan zijn die verschillen veel geringer. Om die netto prestaties vast te kunnen stellen blijkt het nodig om met gegevens op individueel en schoolniveau te werken (aan analyses met geaggregeerde scores kleven namelijk bezwaren, zie hoofdstuk 2).

De verzameling van gegevens op zowel individueel als schoolniveau is gangbaar in wetenschappelijk onderzoek. Aangezien dergelijke gegevens slechts voor een steekproef van scholen en niet voor de gehele populatie beschikbaar zijn, zijn ze niet geschikt voor een onderzoek naar de prestaties van alle scholen. Zo lang deze gegevens er niet voor alle scholen zijn, is een eerlijke vergelijking van scholen onmogelijk.

1.2 Invloeden op prestaties

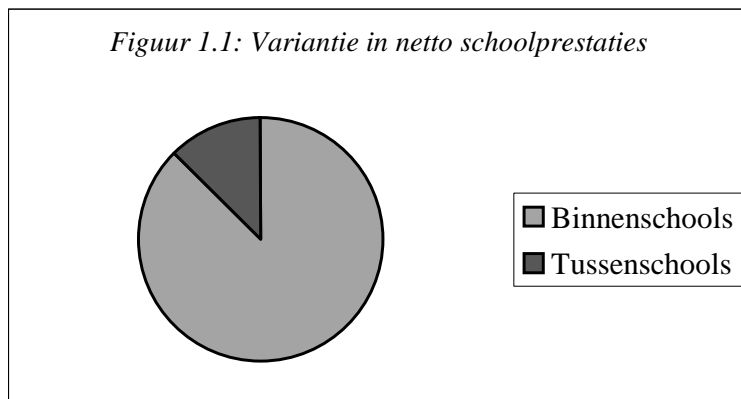
Relatieve binnenschoolse en tussenschoolse variantie

Het Trouw-onderzoek wekt de suggestie dat verschillen in prestaties van leerlingen vooral op het conto van een school moeten worden geschreven. Scholen krijgen van het dagblad namelijk een rapportcijfer en die cijfers variëren sterk. Een kwart van de scholen heeft een vijf of lager als score. De helft van de scholen scoort gemiddeld en heeft een zes als cijfer. Scholen met een zeven of meer maken deel uit van het kwart hoogst scorende scholen. Bij deze rapportcijfers zijn verschillende vragen te stellen, bijvoorbeeld, presteren scholen met een vijf of lager onvoldoende? Dronkers (1997b) vraagt zich dit ook af "De inspectie of overheid heeft nooit vastgesteld bij hoeveel zakkers, zittenblijvers of uitstromers een school onvoldoende functioneert." Het is dan ook mogelijk dat alle scholen in ab-

solute zin naar behoren presteren, maar er toch verschillen tussen scholen bestaan (Goldstein, 1997).

Wat stellen de verschillen in prestaties tussen scholen voor? In de vorige paragraaf heb ik het onderscheid tussen netto en bruto prestaties naar voren gebracht. Uit onderzoek blijkt herhaaldelijk dat verschillen in netto prestaties slechts in geringe mate samenhangen met kenmerken van scholen (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld & York, 1966; Scheerens, 1989). Na correctie voor verschillen in leerlingenpopulatie wijken scholen wat prestaties betreft in beperkte mate van elkaar af.

Om aan te geven in hoeverre scholen in prestaties verschillen is het gemakkelijk om er een bepaalde maat voor te hebben. De maat voor de gemiddelde afwijking van scholen is de *relatieve tussenschoolse variantie*. Deze maat is 0 als verschillen in prestaties helemaal niet aan de school zijn toe te schrijven. Is de school de alles bepalende factor voor verschillen in prestaties, dan is de relatieve tussenschoolse variantie 100 procent. Doorgaans is de relatieve tussenschoolse variantie in netto prestaties 10 tot 15 procent (Scheerens, 1989). Voor de prestaties van leerlingen maakt het dus enigszins uit op welke school ze zitten. Het overgrote deel van de verschillen in prestaties wordt echter door andere kenmerken bepaald, zoals de aanleg van leerlingen of hun gezinsachtergrond. De *relatieve binnenschoolse variantie* is dan ook veel groter dan de relatieve tussenschoolse variantie en bedraagt 85 tot 90 procent. In Figuur 1.1 zijn beide variantiecomponenten grafisch weergegeven. Rapportcijfers voor scholen zouden alleen maar betrekking moeten hebben op de relatieve tussenschoolse variantie.



1.2.1 Leerlingen en gezinnen verschillen

Voor een verklaring van verschillen in prestaties nemen de leerling en het gezin een cruciale positie in. Er zijn dan ook talrijke onderzoekers die de invloed van leerling- en gezinskenmerken op prestaties hebben onderzocht. Veel van dat on-

derzoek gaat over de invloed van structurele kenmerken van leerlingen en gezinnen. Klassiekers op dit terrein zijn internationaal de studies van bijvoorbeeld Blau en Duncan (1967), Duncan, Featherman en Duncan (1972) en Boudon (1974) en in Nederland van Van Heek e.a. (1968) en Peschar (1975). Uit deze onderzoeken komt overduidelijk naar voren dat leerlingen uit lagere sociale milieus gemiddeld genomen slechter presteren in het onderwijs, te lage vervolgoopleidingen kiezen, slechtere banen krijgen en minder gaan verdienen dan leerlingen uit hogere sociale milieus. Dergelijk onderzoek naar onderwijsongelijkheid ging aanvankelijk vooral over verschillen naar sociaal-economische status. In de loop van de tijd is er ook aandacht ontstaan voor de ongelijkheid die kan ontstaan door andere structurele kenmerken, zoals de sekse, de etniciteit en het aantal ouders van een leerling (zie voor overzichten respectievelijk: Dekkers, 1996; Ledoux, 1996; Bosman, 1993).

In veel van het onderzoek naar de invloed van structurele kenmerken op de prestaties van leerlingen valt steeds meer een duidelijke drang te bespeuren om het niet te laten bij de vaststelling dat er sprake is van onderwijsongelijkheid. Meer en meer willen onderzoekers er achterkomen *waarom* er verschillen zijn tussen leerlingen uit lagere en hogere sociale milieus, jongens en meisjes, allochtonen en autochtonen en kinderen uit eenouder- en tweeoudergezinnen. Onderzoekers als Meijnen (1977), De Graaf (1987) en Van der Velden (1991) hebben daarom geprobeerd om empirische ondersteuning te vinden voor theorieën over onderwijsongelijkheid.

In navolging van Kohn (1977²) ging Meijnen (1977) de invloed van sociaal-economische status op gezinsprocessen na. Naarmate het opleidingsniveau van Amerikaanse ouders hoger is, zijn ze volgens Kohn in de *opvoeding* van hun kinderen meer gericht op zelfbepaling en minder op gehoorzaamheid. Kohn stelt dat dit patroon gerelateerd zou zijn aan de werkomstandigheden van ouders. Hoog opgeleide personen werken immers vaak onafhankelijk, kennen geen nauwe supervisie, en hebben minder routinematig en meer complex werk. Daarom zouden ze onafhankelijkheid en zelfontplooiing meer waarderen dan laag opgeleide personen. In Nederland gingen Meijnen (1977) en Van der Velden (1991) na in hoeverre deze theorie hier geldig zou kunnen zijn. Deze onderzoeken bevestigden slechts in beperkte mate de theorie van Kohn.

Een andere theorie ter verklaring van onderwijsongelijkheid is die van Bourdieu (1977) over het *leefpatroon van ouders*. In de opvoeding nemen kinderen volgens Bourdieu allerlei symbolische elementen in hun leefpatroon op. Ze nemen voor een belangrijk deel de smaak en de manieren van hun ouders over, waardoor kinderen de belangstelling voor de 'hogere' cultuur al dan niet van huis uit meekrijgen. Met name de hogere sociale milieus zijn cultureel rijk, want in die kringen lezen per-

sonen meer en gaan ze vaker naar het museum, de schouwburg of het filmhuis. Wanneer ouders een dergelijk leefpatroon hebben, dan is er in het gezin cultureel kapitaal aanwezig. Van dit kapitaal kunnen kinderen bijvoorbeeld profiteren, wanneer ouders activiteiten met hen ondernemen. Naarmate kinderen meer elementen van de hogere cultuur hebben overgenomen, blijken ze verder te komen in het onderwijs en vervolgens ook op de arbeidsmarkt. Docenten of werkgevers waarderen hun fijne smaak, belangstelling voor cultuur en goede manieren. Docenten communiceren gemakkelijker met, besteden meer aandacht aan en denken positiever over dergelijke leerlingen. Verscheidene onderzoekers, waaronder De Graaf (1987) en Van der Velden (1991), hebben laten zien dat er een effect uitgaat van het cultureel kapitaal van ouders op de prestaties van hun kinderen. Bovendien kan dit kenmerk de invloed van de sociaal-economische status ten dele verklaren.

Over de invloed van de sociaal-economische status op prestaties van leerlingen zijn verscheidene theorieën opgesteld. Door talrijke onderzoekers is met wisselend succes geprobeerd om empirisch houvast voor deze theorieën te vinden. Voor de invloed van cultureel kapitaal op prestaties van leerlingen is geregeld bevestiging gevonden. De theorie van Kohn over zelfbepaling en conformiteit is empirisch minder sterk onderbouwd. Een andere grote theorie uit de jaren zeventig, die van Bernstein (1971) over taalgebruik in het gezin en sociale ongelijkheid, is wegens gebrek aan empirische ondersteuning al helemaal uit het zicht verdwenen.

Uit de toepassingen van deze en andere theorieën over de invloed van structurele kenmerken op de loopbaan van leerlingen blijkt dat er nog steeds in beperkte mate een antwoord kan worden gegeven op de vraag waarom er verschillen in prestaties zijn tussen leerlingen. Een van de redenen hiervoor zou kunnen zijn dat de meeste onderzoekers slechts naar de invloed van enkele kenmerken op prestaties kijken. Meijnen (1977) en Van der Velden (1991) hebben bijvoorbeeld alleen de invloed van gezinskenmerken en niet die van leerlingkenmerken op prestaties onderzocht.^{iv} Uit hun onderzoek blijkt dan dat de sociaal-economische status een invloed op de prestaties van leerlingen heeft en dat deze invloed door enkele gezinskenmerken kan worden verklaard. Met name het cultureel kapitaal vormt een belangrijke verklaring voor verschillen in prestaties tussen leerlingen uit lagere en hogere sociale milieus. Waarom het cultureel kapitaal een invloed heeft is echter een vraag die onbeantwoord blijft. Welke mechanismen veroorzaken dat het cultureel kapitaal van ouders een invloed heeft op prestaties van leerlingen? Verloopt de invloed van het leefpatroon van ouders via het leefpatroon van leerlingen? Blijft deze invloed bestaan, wanneer rekening wordt gehouden met de eerdere prestaties van leerlingen? Om dergelijke vragen te beantwoorden, moeten in een

onderzoek gezins- en leerlingkenmerken tegelijkertijd worden opgenomen. Het is daarom nodig om de *black box* van het gezin te openen.

Om die opening te bewerkstelligen is het mijns inziens nodig om een beter denkkader te ontwikkelen. Het uitgangspunt bij het verklaren van verschillen in prestaties zou moeten zijn dat leerlingen en hun gezinsachtergrond een cruciale positie innemen (Wang, Haertel & Walberg, 1993). Verschillen in prestaties kunnen met name worden verklaard door kenmerken op individueel niveau. Op het individuele niveau kan allereerst een onderscheid worden gemaakt tussen *structurele* en *culturele* kenmerken. Voor de culturele kenmerken, zoals het leefpatroon van leerlingen of de opvoedingsstijl in het gezin, geldt dat ze ten dele de invloed van de eerder genoemde structurele kenmerken kunnen intermediairen, zie de onderstaande figuur.



In de bovenstaande figuur valt te zien dat de invloed van structurele kenmerken op prestaties van leerlingen direct en indirect, via culturele kenmerken, kan zijn. Met dit denkkader kan worden nagegaan of de invloed van structurele kenmerken te verklaren valt door culturele kenmerken. Wat deze culturele kenmerken betreft valt een onderscheid te maken tussen kenmerken van leerlingen en gezinnen en tussen kenmerken die gerelateerd zijn aan school en kenmerken die meer algemeen zijn, zie Tabel 1.1 voor enkele voorbeelden van algemene en schoolse kenmerken. Beide dimensies te zamen leiden tot een opdeling van de culturele kenmerken in vier categorieën: algemene gezinskenmerken, schoolse gezinskenmerken, algemene leerlingkenmerken en schoolse leerlingkenmerken, zie het onderstaande schema.

	GEZIN	LEERLING
ALGEMEEN	4	2
SCHOOLS	3	1

Deze vier categorieën zijn te ordenen naar de mate waarin ze direct gerelateerd zijn aan de prestaties van leerlingen. Bij deze ordening is de eerste aanname dat kenmerken van leerlingen meer direct gerelateerd zijn aan hun prestaties dan

kenmerken van hun gezinnen. De tweede aanname voor de culturele kenmerken is dat voor zowel leerling- als gezinskenmerken geldt dat de schoolse kenmerken meer direct zijn gerelateerd aan de prestaties dan de algemene kenmerken. Op deze manier ontstaat dus een ordening van culturele kenmerken, waarbij schoolse kenmerken van leerlingen het meest direct en algemene gezinskenmerken het minst direct aan de prestaties van leerlingen zijn gerelateerd.

De structurele en culturele kenmerken zijn de onafhankelijke variabelen ter verklaring van verschillen in prestaties. Behalve onafhankelijke variabelen kunnen er ook afhankelijke variabelen worden onderscheiden. In dit onderzoek zijn de afhankelijke variabelen de scores op een toets Nederlands (tekstbegrip) en een toets wiskunde.

In totaal onderscheid ik zes categorieën van kenmerken op individueel niveau: één categorie met de afhankelijke variabelen, één categorie met structurele kenmerken en vier categorieën met culturele kenmerken. Met deze zes categorieën is een *structuur van concentrische cirkels* opgesteld, zie Tabel 1.1. De structurele kenmerken zijn in de buitenste cirkel (de vijfde) en de afhankelijke variabelen zijn het centrum van deze structuur geplaatst. De culturele kenmerken zijn als intermediaire variabelen tussen de structurele kenmerken en de afhankelijke variabelen ingeplaatst. De algemene interactie in het gezin, waartoe de opvoedingsstijl kan worden gerekend, vormt cirkel vier. De interactie in het gezin over schoolse zaken vormt cirkel drie. Een voorbeeld hiervan is de controle van huiswerk. Cirkel twee bestaat uit algemene kenmerken van leerlingen, zoals hun tijdsbesteding buiten schooltijd. De schoolse kenmerken van leerlingen vormen cirkel één. Tot de schoolse kenmerken van leerlingen reken ik bijvoorbeeld de eerdere prestaties en de intelligentie van leerlingen.

Met het gebruik van dit kader beoog ik de verschillende invloeden te ordenen en meer inzicht te geven in de (causale) relaties tussen de kenmerken onderling.^v Ook lijkt het mogelijk om met dit kader de effecten van kenmerken als de sociaal-economische status of de etniciteit op leerprestaties te verklaren door intermediaire variabelen. Er wordt geprobeerd om de effecten van kenmerken uit een meer naar buiten gelegen cirkel te verklaren door kenmerken uit een meer naar binnen gelegen cirkel. Hierbij wordt de invloed van kenmerken uit een cirkel gezamenlijk bekeken. Binnen een cirkel wordt er dus niet nader naar een mogelijke ordening van de verbanden gekeken.



Tabel 1.1: Structuur van concentrische cirkels op individueel niveau

	CATEGORIE	KENMERKENDE VARIABELEN
5	structurele kenmerken	sociaal-economische status, sekse, etniciteit, gezinssamenstelling
4	algemene interactie in het gezin	leefpatroon ouders, opvoedingsstijl, aantal contacturen
3	schoolse interactie in het gezin	controle van huiswerk, praten met ouders over school
2	algemene kenmerken van leerlingen	zelfbeeld, leefpatroon (op straat met vrienden, lezen en hobby's, etc.)
1	schoolse kenmerken van leerlingen	eerdere prestaties, intelligentie, inzet, werkhouding, schoolbeleving
0	afhankelijke variabelen	prestaties (Nederlands, wiskunde)

1.2.2 Docenten en scholen verschillen

Sinds het onderzoek van Coleman e.a. (1966) is duidelijk dat het effect van schoolkenmerken klein is in verhouding tot individuele kenmerken. Destijds kwam die bevinding hard aan, omdat vele ouders, onderzoekers en politici hun geloof in de maakbaarheid van de samenleving zagen aangetast. Niet schoolkenmerken maar individuele kenmerken, zoals intelligentie en sociaal milieu, bleken de belangrijkste factoren die verschillen in prestaties bepalen.

Om de achterstand die veel kinderen uit lagere sociale milieus bij de intrede in het onderwijs hebben weg te werken, werden in tal van landen vanaf de jaren zestig programma's opgezet. In de Verenigde Staten waren er landelijke stimuleringprogramma's als Head Start en Follow Through. In Nederland waren er enkele grootschalige programma's, zoals het Innovatieproject Amsterdam (IPA) en Onderwijs en Sociaal Milieu (OSM) te Rotterdam. Over dit laatste project hebben Slavenburg (1986) en Leseman (1990) gerapporteerd. De resultaten van deze compensatieprogramma's zijn veelal teleurstellend. Bij een aantal projecten, zoals bij OSM, zijn er tijdelijke effecten op de prestaties van leerlingen gevonden, maar op de lange termijn worden er geen effecten aangetoond. De evaluaties van deze quasi-experimenten geven zonder uitzondering een weinig opwekkend beeld van de werking van kostbare compensatieprogramma's.^{vi}

Toch zijn er enkele redenen waarom gesteld kan worden dat scholen er wel toe doen. Zo blijken scholen die het ene jaar goed presteren vaak ook het andere jaar goed naar voren te komen. Gray, Jesson, Goldstein, Hedger & Rasbash (1995) heb-

ben in Engeland en Wales de positie van scholen op twee tijdstippen met elkaar vergeleken, zie Tabel 1.2. In deze vergelijking gaat het om de resultaten per school op de General Certificate of Secondary Education. De GCSE is een centraal examen voor verschillende vakken dat wordt afgenomen onder leerlingen van ongeveer zestien jaar. Gray e.a. maken gebruik van netto examenresultaten. Op individueel niveau houden ze rekening met de eerdere prestaties en de sekse van leerlingen. Het blijkt dat ongeveer twee op de drie scholen (68 procent) op beide momenten dezelfde positie inneemt. Ook blijkt dat geen van de scholen die in 1990 hoog scoorde twee jaar later laag scoort, en omgekeerd. De relatieve positie van scholen is dus vrij stabiel.

Tabel 1.2: Netto prestaties van scholen in 1992 in vergelijking met 1990
(Gray e.a., 1995, p. 108)

		1992		
		HOOG	MIDDEN	LAAG
1990	HOOG	15%	6%	0%
	MIDDEN	9%	35%	9%
	LAAG	0%	9%	18%

Verder merken onderzoekers geregeld op dat de praktische betekenis van school-effecten groot is (Bosker & Scheerens, 1989, p. 750): “The most important observation concerning the effect size is, that even small effects may be relevant, if only because the effects should be multiplied by the number of pupils benefiting from outstanding schools.”

In onderzoek naar effecten van schoolkenmerken gaat het tot slot om de vraag of de *kwaliteit* van onderwijs een verschil uitmaakt en niet om de vraag of de *kwantiteit* van onderwijs van belang is. De kwantiteit van scholing blijkt echter belangrijk voor het presteren van leerlingen. In een pleidooi voor het belang van onderwijs wijzen Entwisle, Alexander en Steffel Olson (1997), in navolging van Heyns (1978), erop dat leerlingen in het basisonderwijs tijdens het schooljaar sterk vooruitgaan in prestaties, maar in de zomerperiode weinig bij leren. Dit blijkt het sterkst te gelden voor leerlingen uit lagere sociale milieus. Gedurende het schooljaar blijken leerlingen uit verschillende sociale milieus ongeveer evenveel vooruit te gaan in prestaties. In de zomerperiode boeken leerlingen uit lagere sociale milieus daarentegen minder vooruitgang dan leerlingen uit hogere sociale milieus. Verschillen in prestaties tussen leerlingen naar sociaal milieu ontstaan wanneer leerlingen niet op school zijn, met name in de voorschoolse periode en verder in de zomerperiode (deze invloed blijkt het sterkst in de eerste jaren van

het basisonderwijs). In de periode dat leerlingen wel op school zijn, treden er veel minder verschillen op. De kwantiteit van onderwijs is dus van gewicht voor het presteren van leerlingen. Tijdens de schoolperiode leren leerlingen meer dan in de zomerperiode en de ongelijkheid in prestaties tussen leerlingen uit verschillende sociale milieus neemt niet toe in de periode dat leerlingen op school zitten, maar alleen in de periode dat leerlingen niet op school zijn. Nederlands onderzoek op dit terrein ontbreekt. De verwachting is echter dat deze bevinding ook in andere landen opgang doet.

Effectiviteitsonderzoek

In hoeverre kunnen kenmerken van scholen en docenten verschillen in prestaties verklaren? Op dit terrein zijn er vele onderzoeken verricht. Tot de klassiekers van dit onderzoek behoort het werk van Coleman (1966) en Jencks (1972) met hun collega's. Hoewel in beide studies allereerst de milieubepaaldheid van prestaties en schoolloopbanen wordt geïdentificeerd, kijken deze onderzoekers ook naar de effecten van *materiële schoolkenmerken* (bijvoorbeeld de ouderdom van het schoolgebouw, het aantal boeken in de schoolbibliotheek, het salaris van docenten en de uitgaven per leerling) en enkele *immateriële schoolkenmerken* (houding van schoolleiders en docenten ten opzichte van de leerlingen en het beleid van de school).

Naar de invloed van materiële schoolkenmerken is later vooral door onderwijs economen gekeken. Hieruit blijkt dat er op de prestaties van leerlingen vrijwel geen invloed is van kenmerken als de uitgaven per leerling of het salaris van docenten. Scheerens (1989) noemt de uitkomsten van dit type onderzoek teleurstellend. De uitkomsten van verschillende onderzoeken zijn inconsistent en de effecten van de materiële kenmerken zijn hooguit gering.

De invloed van immateriële schoolkenmerken staat met name centraal in het *schooleffectiviteitsonderzoek*. Onder schooleffectiviteit wordt doorgaans verstaan de component van schoolvorderingen van leerlingen die niet kan worden toegeschreven aan andere invloeden dan die van de school zelf. In schooleffectiviteitsonderzoek gaat het dus om de netto prestaties van scholen. Aangezien de materiële schoolkenmerken weinig of geen invloed hebben op de prestaties van leerlingen, richten schooleffectiviteitsonderzoekers zich meer op de proceskenmerken van scholen. Een van de klassiekers van deze stroming is een pleidooi dat Edmonds hield over de invloed van basisscholen op de prestaties van leerlingen. De kenmerken voor schooleffectiviteit die Edmonds (1979) onderscheidt zijn: (1) de gerichtheid van de schoolleiding op kwaliteit; (2) de hoge verwachtingen bij docenten over de leervorderingen van leerlingen; (3) de ordelijkheid en veiligheid van het schoolklimaat; (4) het accent op de verwerving van basisvaardigheden

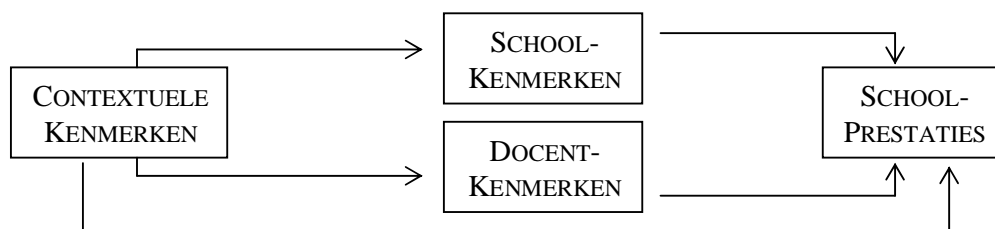
door leerlingen; (5) de frequente evaluatie van de leerlingvoortgang.^{vii} In Nederland geniet het schooleffectiviteitsonderzoek grote belangstelling. Met name Scheerens (1989; 1992) en Creemers (1991; 1994) en hun collega's bij het OCTO en het GION hebben binnen deze onderzoekslijn veel werk verricht.

Behalve naar kenmerken van effectieve scholen kijken onderzoekers soms ook naar kenmerken van effectieve docenten. Deze stroming staat bekend als onderzoek naar *effectieve instructie*. In deze stroming gaat het om verschillen tussen klassen. Creemers (1991, p. 3) stelt dat de praktische betekenis van klassenkenmerken groot is. "Het onderwijsleerproces is een van de weinige mogelijkheden die op de korte termijn en door de leerkracht eventueel alleen is te veranderen in de richting van een grotere effectiviteit." Een groot deel van het onderzoek op dit terrein is onderwijspsychologisch van aard. Om de invloed van klassenkenmerken te bepalen gebruiken onderzoekers vaak observaties of experimenten (Scheerens, 1989). In dit type onderzoek staan kenmerken centraal als gelegenheid tot leren en effectieve leertijd, gestructureerd onderwijs, hoge verwachtingen over prestaties en positieve bekrachtiging.^{viii}

Aangezien de resultaten van grootschalige compensatieprogramma's teleurstellend waren, is in Rotterdam in 1991 het zogeheten Kleinschalig Experiment Achterstandsbestrijding (KEA) gestart. Aan dit project doen slechts vier scholen mee. Uit de evaluatie blijkt dat de leerlingen uit dit KEA-cohort, die op scholen met overwegend allochtone leerlingen zitten, halverwege de basisschool op een zelfde niveau presteren als de gemiddelde leerling van andere basisscholen. Dit succes schrijven de onderzoekers (Hoogendijk & Wolfgram, 1995) toe aan een klassenkenmerk, namelijk de nadruk op gestructureerd onderwijs door leerkrachten. De resultaten van het KEA-onderzoek vormen dus een bevestiging van het belang van effectieve instructie.

Tot slot zijn er onderzoeken waarin vooral *contextuele schoolkenmerken* centraal staan, zie Tabel 1.3 voor enkele voorbeelden van dergelijke kenmerken. Aan de basis van deze stroming staat een onderzoek van Coleman, Hoffer en Kilgore (1982), waarin wordt geconcludeerd dat Amerikaanse scholen in de particuliere sector effectiever zijn dan scholen in de publieke sector. Bovendien stellen zij dat katholieke scholen, de overgrote meerderheid van de particuliere sector in de Verenigde Staten, bijzonder succesvol zijn in het vergroten van de onderwijskansen van leerlingen uit achterstandsgroepen. Katholieke scholen beschikken over een compenserend vermogen voor leerlingen uit lagere sociale milieus. Als redenen voor dit succes, dragen zij aan dat particuliere scholen meer nadruk op basisvaardigheden leggen en dat het klimaat op dergelijke scholen geschikter is voor het leren. Op katholieke scholen gedragen leerlingen zich positiever en tonen de schoolleiding en docenten meer belangstelling voor leerlingen. Ook latere studies

van Coleman en Hoffer (1987) en Bryk, Lee en Holland (1993) wijzen op het succes van katholieke scholen in de Verenigde Staten. In Nederland levert het onderzoek naar verschillen tussen openbare en bijzondere scholen een minder eenduidig beeld op (Dijkstra, 1992; 1997).



Er zijn drie soorten kenmerken op klas- en schoolniveau: docent-, school- en contextuele kenmerken. In verschillende onderzoeken naar effectiviteit is gewezen op het belang van dit soort kenmerken. Zoals op individueel niveau de effecten van structurele kenmerken deels te verklaren zijn door culturele kenmerken, kunnen contextuele effecten mogelijk worden verklaard door docent- en schoolkenmerken, zie de bovenstaande figuur. In het onderzoek naar verschillen tussen openbare en bijzondere scholen wordt op die manier dus de *black box* van de school geopend. Niet het label katholiek vormt de verklaring voor het succes van katholieke scholen in de Verenigde Staten, maar bijvoorbeeld de nadruk op basisvaardigheden of het schoolklimaat. In dergelijk verklarend onderzoek is het de bedoeling om de invloed van contextuele kenmerken, zoals de denominatie of de locatiegrootte, zo veel mogelijk te laten intermediairen door docent- en schoolkenmerken.^{ix}

Tabel 1.3: Structuur van concentrische cirkels op klas- en schoolniveau

	CATEGORIE	KENMERKENDE VARIABLEN
C	contextuele kenmerken	denominatie, percentage allochtone leerlingen, locatiegrootte
B	schoolkenmerken	nadruk op basisvaardigheden, orde- lijkheid van het schoolklimaat
A	docentkenmerken	gelegenheid tot leren, gestructureerd onderwijs, positieve bekrachtiging

1.3 Probleemstelling

Afbakening

In dit onderzoek vormen prestaties en vorderingen van leerlingen de rode draad. Die prestaties en vorderingen probeer ik te verklaren door te kijken naar kenmerken van enerzijds leerlingen en gezinnen en anderzijds scholen en docenten. Hierbij kijk ik dus tegelijkertijd naar kenmerken op verschillende niveaus.

Voor dit onderzoek heb ik enkele keuzes gemaakt. Allereerst heeft deze studie betrekking op *Nederland*. Voor de invloed van leerlingen en gezinnen geldt dat de overeenkomsten tussen onderzoeken in verschillende landen groot zijn. Voor de invloed van scholen en docenten geldt dat minder (Scheerens & Bosker, 1997). Eerdere studies in andere landen vormen geen empirische bewijsvoering voor effecten van bepaalde kenmerken in Nederland. Omgekeerd zijn de uitkomsten van dit onderzoek dan ook niet gelijk naar andere landen te vertalen.

In de tweede plaats richt ik me in deze studie op de *onderbouw van het voortgezet onderwijs*, de periode van de basisvorming. Onderzoek binnen een bepaald type van onderwijs valt niet zonder meer over te zetten naar andere typen van onderwijs. Sommige kenmerken van leerlingen, gezinnen, scholen of docenten hebben bijvoorbeeld een grotere invloed in het voortgezet onderwijs dan in het basis-, middelbaar of hoger onderwijs (Scheerens & Bosker, 1997).

In de derde plaats ga ik de invloeden na van verschillende kenmerken op de toetsscores van leerlingen op de vakken *Nederlands* (tekstbegrip) en *wiskunde*. Cognitieve prestaties gelden doorgaans als de belangrijkste graadmeter voor een vergelijking van leerlingen of scholen. Voor de onderbouw van het voortgezet onderwijs zijn de scores op Nederlands en wiskunde de meest geschikte maten die voorhanden waren.

In de vierde plaats heb ik in dit onderzoek gebruikt gemaakt van *schriftelijke vragenlijsten*. Deze manier van gegevensverzameling past goed bij een uitgebreid onderzoek, waarbij gegevens van leerlingen, ouders, docenten en schoolleiders nodig zijn. Het nadeel is echter dat aan met name kenmerken van docenten op deze manier waarschijnlijk te kort wordt gedaan. Uit onderzoek blijkt dat gegevens over lesgeven beter kunnen worden achterhaald door observaties.

Deze studie is *kwantitatief* van aard. In een eerder rapport (Veenstra, 1997) heb ik acht scholen voor voortgezet onderwijs kwalitatief beschreven. Door die scholen te bezoeken heb ik meer inzicht gekregen in het voortgezet onderwijs, ben ik te weten gekomen dat de directiegegevens uit het cohortonderzoek VOCL'93 (Voortgezet Onderwijs Cohort Leerlingen) in grote lijn een goede weerspiegeling van de praktijk geven en heb ik nieuwe, schriftelijke vragen voor leerlingen, do-

centen en directieleden uitgeprobeerd. In deze studie bouw ik voort op dat kwalitatieve onderzoek. Nu gebruik ik gegevens van 150 scholen.

Ondanks deze afbakening, is de opzet van dit onderzoek breed. Ik behandel niet de afzonderlijke invloed van een bepaald kenmerk, bijvoorbeeld de opvoedingsstijl van ouders, de inrichting van het mentoraat of de invoering van de basisvorming, op de prestaties van leerlingen, maar de invloed van talrijke kenmerken *tegelijkertijd*.

Bij de beantwoording van een bepaalde vraag spelen vele theoretische inzichten en oriëntaties op de achtergrond een rol mee. Door deze grootscheepse aanpak gaan nuances soms misschien verloren. Daar staat echter tegenover dat voor de onderbouw van het Nederlandse voortgezet onderwijs een integrale studie naar de verschillende invloeden op prestaties en vorderingen ontbreekt. Hoewel niet vanuit één theorie wordt gewerkt, beschouw ik dit onderzoek wel als theoretisch relevant. Door de integrale aanpak wordt duidelijk hoe theorieën en onderzoeken op verschillende deelterreinen zich tot elkaar verhouden. Om de brede aanpak inzichtelijk te houden, heb ik ervoor gekozen om de kenmerken op individueel, klas- en schoolniveau te ordenen in een structuur van concentrische cirkels (Tabel 1.1 en 1.3).

Probleemstelling

Deze studie is een kwantitatief onderzoek naar de prestaties en vorderingen van leerlingen in de onderbouw van het voortgezet onderwijs van Nederland. Aansluitend bij de discussie over openbare prestatiegegevens van scholen, wil ik allereerst nagaan aan welke eisen een dergelijke publieke verantwoording moet voldoen. Hierbij wil ik met name nagaan in hoeverre de schattingen van de netto prestaties van scholen verschillen in een analyse met geaggregeerde gegevens en een multiniveau analyse.

In de tweede plaats wil ik met multiniveau analyse achterhalen welke leerling- en gezinskenmerken van invloed zijn op de prestaties van leerlingen. Ook wil ik nagaan of docent- en schoolkenmerken van invloed zijn op de prestaties van leerlingen. Centraal staat de vraag in hoeverre de prestaties van leerlingen op één bepaald tijdstip kunnen worden verklaard uit allerlei kenmerken van leerlingen, gezinnen, docenten en scholen. Om de netto prestaties van leerlingen te bepalen wordt hierbij rekening gehouden met de prestaties van leerlingen op een eerder tijdstip. De toetsscores in leerjaar drie zijn de afhankelijke variabelen en de eerdere prestaties in leerjaar één behoren tot de onafhankelijke variabelen. Een dergelijk model is standaard in het schooleffectiviteitsonderzoek en staat bekend als het *covariantie-analytisch model*.

Met het covariantie-analytisch model wordt naar invloeden op prestaties (toestand) en niet naar invloeden op vorderingen (verandering) gekeken. Met dit model kan dan ook niet over effecten op vooruitgang in prestaties worden gesproken. Om naar vooruitgang in prestaties te kijken, moet voor een *variantie-analytisch model van verschillen in (relatieve) leerwinst* worden gekozen. Als laatste stap wil ik daarom de invloed van diverse kenmerken op de vorderingen van leerlingen onderzoeken. Mijn probleemstelling bestaat uit drie vragen, waarbij de laatste twee zijn onderverdeeld:

- 1 In hoeverre verschillen de schattingen van de netto prestaties van scholen in een analyse met geaggregeerde gegevens en een multiniveau analyse?
- 2a In hoeverre en waarom zijn kenmerken op *individueel niveau* van invloed op de *prestaties* van leerlingen in het derde leerjaar van het voortgezet onderwijs?
- 2b In hoeverre en waarom zijn kenmerken op *klas- en schoolniveau* van invloed op de *prestaties* van leerlingen in het derde leerjaar van het voortgezet onderwijs?
- 3a In hoeverre en waarom zijn kenmerken op *individueel niveau* van invloed op de *vorderingen* (verschillen in relatieve leerwinst) van leerlingen tussen het eerste en het derde leerjaar van het voortgezet onderwijs?
- 3b In hoeverre en waarom zijn kenmerken op *klas- en schoolniveau* van invloed op de *vorderingen* (verschillen in relatieve leerwinst) van leerlingen tussen het eerste en het derde leerjaar van het voortgezet onderwijs?

Relevantie

Schooldirecties, docenten, ouders en leerlingen hebben allen in meer of mindere mate een idee hoe deze prestaties en vorderingen worden beïnvloed. Door onderzoek naar de effecten hierop kunnen hun ideeën mogelijk aan kracht winnen of inboeten. Om dezelfde reden worden ook politici en onderwijsambtenaren mogelijk tot denken gezet door dit onderzoek. Zij kunnen wellicht ideeën opdoen of steun vinden voor hun beleid over bijvoorbeeld het verminderen van de ongelijkheid in onderwijskansen of het verhogen van de kwaliteit van het onderwijs.

Wetenschappelijk gezien is het doel van dit onderzoek viervoudig. Om verschillen in prestaties en vorderingen tussen leerlingen te verklaren wordt in dit onderzoek een denkkader beproefd, de *structuur van concentrische cirkels*. Met deze structuur wordt gepoogd de invloeden op prestaties en vorderingen meer te ontrafelen. Hierbij wordt in samenhang gekeken naar de invloed van leerling-, gezins-, klas- en schoolkenmerken. Voor de bestudering van deze verschillende kenmerken beschik ik over gegevens van schooldirecties, docenten, leerlingen en ouders. In de tweede plaats laat ik zien dat de analyses voor *alle schooltypen tege-*

lijktijd kunnen worden gedaan. Door interacties op te nemen blijkt het niet nodig om gegevens uit te splitsen per schooltype. Ten derde kijk ik niet alleen naar de invloeden op prestaties (met een covariantie-benadering), maar ook naar de invloeden op de relatieve leerwinst (met een verschilscore-benadering). Op deze manier kan er worden gekeken naar de *voor- of achteruitgang in prestaties* van leerlingen. Uit een vergelijking van de twee analysemethoden zal blijken dat beide methoden elkaar goed aanvullen. Tot slot probeer ik in dit onderzoek vooruitgang te boeken door uitgebreid van de mogelijkheden van *multiniveau analyse* gebruik te maken. Ik hanteer een geneste structuur met drie niveaus: leerlingen binnen klassen binnen scholen. Verder analyseer ik de invloed van de verschillende kenmerken op de toetsscores Nederlands en wiskunde multivariaat, dat wil zeggen de analyses vinden in één keer plaats voor de twee afhankelijke variabelen, Nederlands en wiskunde. In het volgende hoofdstuk bespreek ik de meerwaarde van multiniveau analyse en een multivariate aanpak.

De invalshoek die ik in dit boek heb gekozen is een combinatie van verschillende sociale wetenschappen (plus statistiek). Dit onderzoek sluit aan bij de onderwijskunde, omdat het verwantschap vertoont met het schooleffectiviteitsonderzoek (Scheerens & Bosker, 1997). In effectiviteitsonderzoek gaat de belangstelling primair uit naar kenmerken van scholen of docenten. Van de kenmerken op individueel niveau komen vooral structurele kenmerken aan bod, met name dan de sekse, de sociaal-economische status en de etniciteit. Door ook te kijken naar culturele kenmerken van leerlingen en gezinnen sluit dit onderzoek ook aan bij psychologisch en sociologisch onderzoek.

Opzet

In de voorgaande paragrafen zijn de belangrijkste lijnen voor dit boek al uiteengezet. In de volgende hoofdstukken bouw ik hier op voort. In hoofdstuk 2 maak ik duidelijk dat om de netto prestatie van een school vast te kunnen stellen het nodig is om gegevens op individueel en schoolniveau te verzamelen. In dat hoofdstuk bespreek ik allereerst de beginselen van multiniveau analyse. Het belang van deze methode illustreer ik met het Trouw-onderzoek naar schoolprestaties.

In hoofdstuk 3 ga ik in op de vraag welke leerling- en gezinskenmerken van invloed zijn op de prestaties van leerlingen. Om deze vraag te beantwoorden hanteer ik de structuur van concentrische cirkels. De resultaten van de multiniveau analyse die gebaseerd zijn op dit kader staan centraal in dat hoofdstuk. De vraag naar de invloed van school- en docentkenmerken komt aan bod in hoofdstuk 4.

Voor de analyses uit hoofdstuk 3, 4 en 5 gebruik ik gegevens afkomstig uit het cohortonderzoek VOCL'93. Deze gegevens zijn verzameld door het Centraal Bureau voor de Statistiek (CBS), het Gronings Instituut voor Onderwijsonderzoek

(GION) en het Onderzoek Centrum Toegepaste Onderwijskunde (OCTO). Voor de analyses verricht ik multiniveau analyse, met informatie van 7.000 leerlingen, afkomstig uit 450 klassen van 150 scholen. Deze leerlingen zitten op het ivbo, het vbo, de mavo, de havo of het vwo en zijn in het schooljaar 1993-'94 aan het voortgezet onderwijs begonnen.

In hoofdstuk 5 vergelijk ik twee modellen voor het analyseren van prestaties van leerlingen. Aan de ene kant hanteer ik een multivariaat *covariantie*-analytisch model met de prestaties van leerlingen in het derde leerjaar als afhankelijke variabelen en het beginniveau als covariaat. Aan de andere kant stel ik een multivariaat variantie-analytisch model voor de analyse van *relatieve leerwinst* op, waarbij de verschilscore tussen eind- en beginniveau de afhankelijke variabele is. In hoofdstuk 6 vat ik de uitkomsten van dit onderzoek samen, geef ik de reikwijdte van het onderzoek weer en ga ik in op vragen voor vervolgonderzoek. In hoofdstuk 2 tot en met 5 worden in verscheidene tabellen onderzoeksresultaten weergegeven. Een deel van de tabellen is niet per se nodig voor het betoog. Deze tabellen zijn als achtergrond in een bijlage opgenomen.

Eindnoten

ⁱ Niet alleen Trouw maar ook andere dagbladen, zoals de Volkskrant, en weekbladen, bijvoorbeeld Elsevier en HP/De Tijd, publiceren geregeld hitlijsten van scholen. Voor het voortgezet en hoger onderwijs zijn er verder keuzegidsen uitgebracht (voor het voortgezet onderwijs: Agerbeek, 1999a).

ⁱⁱ Deze publicaties van schoolgegevens zijn in Engeland te raadplegen via internet: <http://www.open.gov.uk/dfce/perform.htm>.

ⁱⁱⁱ Deze kaarten zijn te raadplegen via internet: <http://kwaliteitskaart.owinsp.nl/kwaliteit/>.

^{iv} De Graaf (1987) kijkt ook voornamelijk naar de invloed van gezinskenmerken (sociaal-economische status, financieel en cultureel kapitaal van ouders) op de schoolloopbaan van leerlingen. In één van zijn analyses neemt hij ook een leerlingkenmerk op, namelijk het leefpatroon van leerlingen. Juist in die analyses ontbreken de eerdere prestaties van leerlingen.

^v Voor de analyse van (causale) relaties zijn padmodellen of LISREL-modellen uiteraard ook geschikt. Wanneer in een onderzoek echter de relaties tussen enkele tientallen variabelen worden onderzocht, leiden die modellen al snel tot onontwarbare patronen van relaties. Het gebruik van een structuur van concentrische cirkels zou in dat soort gevallen een alternatief kunnen vormen.

^{vi} Toch beschouwen Zigler en Muenchow (1992) het programma Head Start als succesvol. Zij wijzen erop dat dit programma langetermijneffecten heeft, zoals hogere diploma's, een grotere arbeidsdeelname en minder criminaliteit van de voormalige deelnemers. Ook scoren ze soms goed wat bijvoorbeeld zelfbeeld of motivatie betreft. De bewijsvoering voor deze successen is echter zwak, omdat Head Start formeel nauwelijks is geëvalueerd. Bij gebrek aan beter halen Zigler en Muenchow, die beiden betrokken waren bij het project, veelvuldig de resultaten van vergelijkbare studies aan. Van een objectieve analyse is in dit geval echter geen sprake.

^{vii} Edmonds (1979, p. 22) haalt zes kenmerken in zijn artikel aan. De betekenis van een van die kenmerken is onduidelijk, te weten: "When necessary, school energy and resources can be diverted

from other business in furtherance of the fundamental objectives". In navolging van bijna alle onderzoekers die Edmonds aanhalen, laat ik dit kenmerk achterwege. Zie verder Scott en Walberg (1979) voor een kritiek op Edmonds.

^{viii} Van structurele kenmerken van docenten, zoals hun opleiding en ervaring, gaat volgens Creemers (1991) een gering of geen effect uit op de prestaties van leerlingen. De invloed van onderwijsleermateriaal op de resultaten van leerlingen blijkt moeilijk aan te tonen. Een mogelijke verklaring hiervoor is de uiteenlopende manier waarop docenten met het materiaal omgaan. Verder zou kunnen gelden dat de meeste methoden hun waarde hebben en dat het er vooral omgaat of docenten weten wanneer ze wat kunnen gebruiken.

^{ix} In de figuur uit paragraaf 1.2.2 is geen pijl getekend tussen het docent- en schoolniveau. Vanuit verschillende analyseconcepten, zoals het *indirecte* of het *recursieve* analyseconcept (Scheerens & Bosker, 1997), zou een dergelijke relatie te beargumenteren zijn. Het zou zo kunnen zijn dat schoolkenmerken niet direct van invloed zijn op schoolprestaties, maar via docentkenmerken invloed hebben (indirect) of dat docent- en schoolkenmerken elkaar wederzijds beïnvloeden (recursief). In dit onderzoek wordt voor docent- en schoolkenmerken echter een *additief* analyseconcept gehanteerd.

2 | Een geneste structuur

2.1 Inleiding

Om verschillen in prestaties van leerlingen te verklaren zijn kenmerken op verschillende niveaus nodig. Deze verschillende niveaus worden in dit onderzoek in een geneste structuur weergegeven, namelijk leerlingen binnen klassen binnen scholen.

Op soortgelijke wijze kan voor de verklaring van verschillen in ziekteverzuim van docenten bijvoorbeeld worden gekeken naar kenmerken op verschillende niveaus. Voor de verklaring zijn bedrijfskenmerken, zoals het schooltype, het personeelsbeleid of de gemiddelde leeftijd van het docententeam, niet genoeg. Een deel van de verschillen kan namelijk worden toegeschreven aan persoonskenmerken, zoals de sekse, de leeftijd, het aantal lesuren of de motivatie van docenten. Om de vraag naar het ziekteverzuim van docenten te beantwoorden kan dus het beste een geneste structuur worden gehanteerd, namelijk docenten genest binnen scholen. In onderzoek naar verschillen in het gemiddeld aantal dagen dat artsen patiënten na een bepaalde operatie in het ziekenhuis laten verblijven, is de multiniveau structuur patiënten genest binnen artsen binnen ziekenhuizen. Om te achterhalen in hoeverre het loon van werknemers verschilt per bedrijfstak en per land is een structuur op te stellen met werknemers genest binnen bedrijfstakken binnen landen. Voor talrijke vraagstukken in de sociale wetenschappen is het van belang om met de hiërarchie van gegevens rekening te houden.

Door de ontwikkeling van technieken voor multiniveau analyse is het mogelijk geworden om data afkomstig van geneste structuren te analyseren.ⁱ In de jaren tachtig kwam deze ontwikkeling in een stroomversnelling door het werk van onder andere Aitkin en Longford (1986), Goldstein (1995²) en Raudenbush en Bryk (1986). Tot die tijd analyseerden wetenschappers hiërarchische gegevens gewoonlijk op één niveau. Om op één niveau te analyseren, moest er een keuze gemaakt worden tussen *desaggregeren* of *aggregeren*. Aan beide methoden zijn echter grote nadelen verbonden. Bij desaggregatie worden gegevens van een hoger op een lager niveau geanalyseerd. In een onderzoek met gegevens van 7.000 leerlingen en 150 scholen, worden de schoolgegevens dan op individueel niveau geanalyseerd. Het aantal onderzoekselementen op schoolniveau wordt zo

op miraculeuze wijze vermenigvuldigd, want er wordt gedaan alsof er over leerlingen en scholen evenveel verschillende, onafhankelijke gegevens zijn. De effecten van gegevens op een hoger niveau worden door desaggregatie opgeblazen.

Door individuele gegevens naar een hoger niveau te aggregeren kunnen een verlies aan onderscheidingsvermogen (*power*), ecologische fouten en interpretatieproblemen ontstaan. In het voorbeeld met 7.000 leerlingen en 150 scholen wordt bij aggregatie het aantal onderzoekselementen teruggebracht tot 150, waardoor het moeilijker wordt om een onderscheid te maken tussen verbanden die wel en die niet ‘waar’ zijn. Door de verkleining van het aantal onderzoekseenheden neemt de kans op de zogeheten *fout van de tweede soort* toe. Dit is de kans op het ten onrechte handhaven van de nulhypothese. Ecologische fouten en interpretatieproblemen kunnen voorkomen doordat individuele gegevens op hoger niveau worden geanalyseerd. Het klassieke voorbeeld van een ecologische fout is de relatie tussen het percentage zwarten en het stemgedrag in Amerikaanse wijken (Robinson, 1950). Naarmate er meer zwarten in een wijk wonen, wordt er meer op de Republikeinse partij gestemd. Uit dit verband zou de verkeerde afleiding kunnen worden gemaakt dat zwarten vaak Republikeins stemmen. Met gegevens op wijkniveau kunnen echter geen uitspraken op individueel niveau worden gedaan. Uit deze geaggregeerde gegevens kan dus niet worden opgemaakt wat blanken of zwarten stemmen.

Een voorbeeld van een interpretatieprobleem kan worden toegelicht met het Trouw-onderzoek. Uit dat onderzoek blijkt dat er een verband is tussen brede scholen en het rendement van de havo en vwo. Brede scholen blijken goed uit te pakken voor de hogere schooltypen van voortgezet onderwijs. De vraag die een ouder zich zou willen stellen is echter of brede scholen ook goed uitpakken voor potentiële havo- en vwo-leerlingen. Die vraag kan niet met de gegevens van Trouw worden beantwoord.ⁱⁱ Ook kan niet worden nagegaan waarom scholen met veel allochtone leerlingen nu slechter presteren in het Trouw-onderzoek. Komen deze slechtere prestaties doordat allochtone leerlingen (een kenmerk op individueel niveau) een achterstand hebben? Of komt dit doordat het percentage allochtone leerlingen (een kenmerk op schoolniveau) van invloed is op prestaties?

Voor de analyse van geneste structuren is multiniveau analyse dus aangewezen (Bryk & Raudenbush, 1992; Goldstein, 1995²; Hox, 1994; Snijders & Bosker, 1999; Woodhouse, Rasbash, Goldstein & Yang, 1996). Multiniveau analyse heeft een aantal voordelen. Door rekening te houden met de hiërarchie van gegevens zijn de standaardfouten, betrouwbaarheidsintervallen en significantietoetsen nauwkeuriger. Hierbij geldt dat deze statistieken voor een analyse van een geneste

structuur doorgaans ‘conservatiever’ zijn in een multiniveau analyse dan in een gedesaggregeerde analyse. In een gedesaggregeerde analyse zijn de standaardfouten vaak kleiner dan in een multiniveau analyse. Als de regressiecoëfficiënten in beide methoden even groot zijn, dan is het significantieniveau van een effect in een analyse op één niveau lager dan in een analyse op meer niveaus. Het significantieniveau wordt immers bepaald door de verhouding tussen de regressiecoëfficiënt en de standaardfout. Deze sterkere significantie is echter schijn, omdat hij berust op de onjuiste assumptie van onafhankelijke onderzoekseenheden (in dit geval leerlingen).

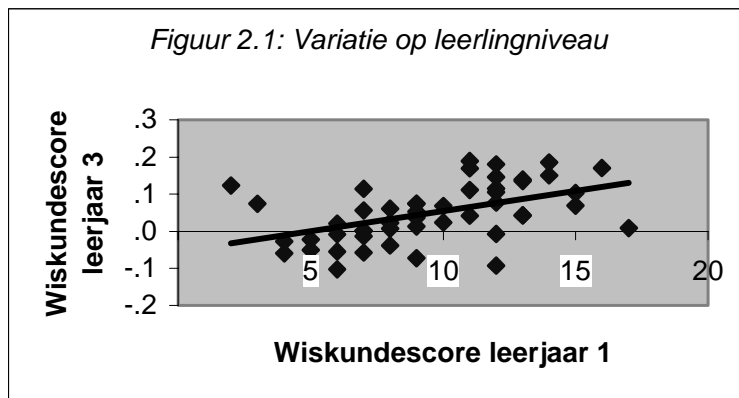
In de tweede plaats geldt dat er met multiniveau analyse onderscheid gemaakt kan worden tussen regressie-effecten binnen en tussen scholen, waardoor er minder interpretatieproblemen optreden. De vraag of etniciteit niet alleen op individueel maar ook op schoolniveau invloed heeft op prestaties, kan met multiniveau analyse worden beantwoord. In de derde plaats kan met multiniveau analyse naar differentiële effecten worden gekeken: is een school voor alle leerlingen even goed? Er kan bijvoorbeeld worden nagegaan of scholen meer verschil uitmaken voor leerlingen uit lagere sociale milieus dan voor leerlingen uit hogere sociale milieus. Ten vierde geldt dat het met multiniveau analyse eenvoudig is om scholen te rangordenen naar hun netto prestatie. Hiervoor kan gebruik worden gemaakt van wat genoemd wordt *posterior means*. In paragraaf 2.3 worden de scholen op deze manier geordend. Verder geldt dat met multiniveau analyse veelal recht wordt gedaan aan de hiërarchische structuur van de *steekproefselectie*. Zo is de steekproef van dit onderzoek in drie stappen geselecteerd. Eerst zijn er scholen getrokken, vervolgens klassen en uit die klassen zijn alle leerlingen genomen. In dit onderzoek is er dus sprake van een drietraps-steekproef.

Voor de analyse van geneste structuren is multiniveau analyse dus noodzakelijk en kan er niet worden volstaan met eenvoudigere, maar ongeschiktere analyses op één niveau. In paragraaf 2.2 van dit hoofdstuk behandel ik de basis van multiniveau analyse. Deze paragraaf biedt de technische achtergrond voor de komende hoofdstukken. In paragraaf 2.3 laat ik zien dat het voor een oordeel over de prestaties van scholen veel uitmaakt of er op één of meer niveaus wordt geanalyseerd.

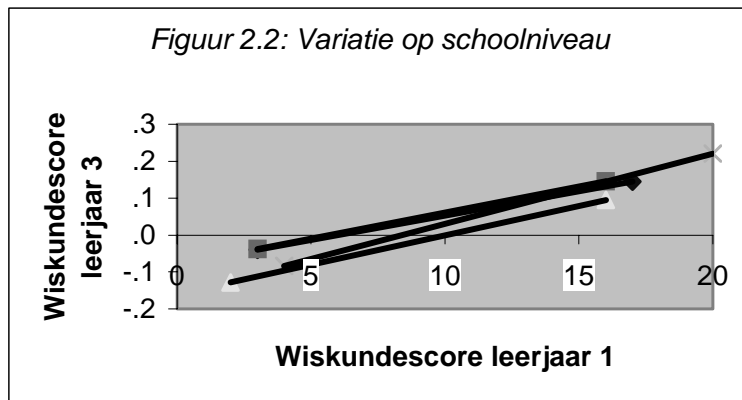
2.2 Multiniveau analyse

Verschillende niveaus

Multiniveau analyse zal ik in de komende hoofdstukken gebruiken voor het verklaren van de invloed van school-, docent-, leerling- en gezinskenmerken op de prestaties en vorderingen van leerlingen. In deze analyses maak ik gebruik van vier niveaus. Niveau nul is een *pseudo-niveau*, dat nodig is voor een analyse met verschillende afhankelijke variabelen tegelijk. Op dat niveau bevinden zich de scores op de toetsen Nederlands en wiskunde. Het eerste niveau is het *individuele niveau* met leerling- en gezinskenmerken. Op het tweede en derde niveau, het *klas-* en *schoolniveau*, zijn de docent- en schoolkenmerken gemodelleerd. De multiniveau structuur voor dit onderzoek is dus afhankelijke variabelen genest binnen leerlingen binnen klassen binnen scholen.



In multiniveau analyse gaat het om het vinden van een formule waarmee de scores op de afhankelijke variabelen zo goed mogelijk geschat kunnen worden vanuit de scores op de onafhankelijke variabelen. In Figuur 2.1 is elke score als een ruit weergegeven. Om in een lineaire regressieanalyse op één niveau de afhankelijke variabele Y te voorspellen uit de onafhankelijke variabele X is het nodig om een vergelijking van een rechte lijn op te stellen, waarbij Y wordt voorspeld uit het *intercept*, de waarde van Y als X gelijk is aan nul, en de *helling* (of: regressiecoëfficiënt; in het Engels: *slope*) van de lijn met de X -as. Wanneer de voorspelling van Y uit X niet perfect is, dan liggen niet alle scores van Y op de regressielijn, zoals in Figuur 2.1 het geval is. De afwijkingen van deze scores tot de regressielijn is de variatie die niet verklaard is, de onverklaarde of residuele *variantie*. Tussen de scores op Y en de voorspellingen van Y door X zijn er dan verschillen. Deze verschillen worden aangeduid met de term *residu* (of: fout), omdat het dat gedeelte van de score van Y is dat niet wordt voorspeld door de regressievergelijking.



Het multiniveau model bouwt voor een deel voort op het lineaire regressiemodel. In Figuur 2.2 is voor vier scholen een regressielijn weergegeven waarin de relatie tussen de wiskundetoetsen in leerjaar één en drie wordt getoond. Omwille van de presentatie ontbreken de scores van leerlingen in Figuur 2.2. Voor alle vier de scholen is de relatie tussen de beide toetsen positief. Naarmate leerlingen in het eerste leerjaar beter scoren op wiskunde behalen ze twee jaar later gemiddeld genomen een hogere score op dit vak. De regressielijnen zijn echter niet identiek. Scholen verschillen namelijk in intercept en in helling. Het intercept geeft de score aan die leerlingen in leerjaar drie behalen, wanneer ze in leerjaar één nul punten scoorden op de wiskundetoets. In dit voorbeeld ligt het punt $x=0$ niet in het bereik van x . Een verschil in helling houdt in dat de relatie tussen de scores in beide leerjaren van school tot school verschilt. Op scholen waar de relatie sterker (een steilere helling) is, gaan de prestaties van leerlingen meer uiteenlopen. Deze scholen worden dan ook wel *selectief* genoemd. Op scholen waar de relatie zwakker (een vlakkere helling) is, treedt een nivellering in prestaties van leerlingen op. Deze scholen worden daarom *egalitair* genoemd. Met verschillen in zowel het intercept als de helling kan de variatie tussen scholen complex zijn.

Multiniveau modellen

Het meest eenvoudige model voor een multiniveau analyse met drie niveaus en één afhankelijke variabele is het *lege model*. Het lege model voor de verklaring van Y bestaat uit de volgende vier termen: het intercept en de residuen op school-, klas- en individueel niveau.

$$(2.1) Y_{ijk} = \beta_{000} + V_{00k} + U_{0jk} + R_{ijk}$$

In het model is i de index voor leerlingen, j de index voor klassen en k de index voor scholen. Wanneer een variabele de index ijk heeft dan varieert de score op leerling-, klas- en schoolniveau. Wanneer een variabele de index $00k$ heeft dan varieert het alleen op schoolniveau. In de vergelijking zijn V_{00k} , U_{0jk} en R_{ijk} random variabelen met een gemiddelde van nul en, zo wordt aangenomen, een normale verdeling en respectievelijk als residuele variantie Φ_v^2 , Φ_u^2 en Φ_r^2 . Door de opname van de *random effects* (V_{00k} , U_{0jk} en R_{ijk}) in de analysestructuur wordt multiniveau analyse gekenmerkt.

Het lege model is nuttig om de residuele varianties tussen scholen, klassen en leerlingen uit te rekenen. In hoofdstuk 1 werd in dit verband gesproken over relatieve binnenschoolse en tussenschoolse variantie. Hierbij is de *intra-klasse correlatiecoëfficiënt* de maat om de verhouding van de residuele varianties aan te geven. De intra-klasse correlatiecoëfficiënt voor het klas- en schoolniveau ten opzichte van het individueel niveau is de variantie *tussen* scholen en klassen gedeeld door de totale variantie: $(\Phi_v^2 + \Phi_u^2) / (\Phi_v^2 + \Phi_u^2 + \Phi_r^2)$. De intra-klasse correlatiecoëfficiënt voor het schoolniveau ten opzichte van het klasniveau is: $\Phi_v^2 / (\Phi_v^2 + \Phi_u^2)$. De intra-klasse correlatiecoëfficiënt is dus de proportie variantie die op hoger niveau wordt aangetroffen. Het is een correlatiecoëfficiënt, omdat deze gelijk is aan de correlatie tussen de scores van twee random getrokken onderzoekselementen (micro-units) in dezelfde, random getrokken macro-unit. In de volgende hoofdstukken wordt deze correlatiecoëfficiënt veelvuldig gehanteerd en zal de formule van voorbeelden worden voorzien.

De tweede stap is om het lege model uit te breiden door *vaste effecten* van onafhankelijke variabelen op het *individuele niveau* op te nemen, bijvoorbeeld de intelligentie van een leerling. Door de opname van deze variabele wordt getracht om de variatie in Y te verklaren. Met één onafhankelijke variabele X ziet het model er als volgt uit:

$$(2.2) Y_{ijk} = \beta_{000} + \beta_{100} x_{ijk} + V_{00k} + U_{0jk} + R_{ijk}$$

Behalve uit een intercept en drie residuele variantiecomponenten bestaat het model uit een regressiecoëfficiënt β_{100} . Deze coëfficiënt kan op de gebruikelijke manier worden geïnterpreteerd: een toename van een eenheid X leidt tot een toename in Y van β_{100} eenheden. Het intercept en de regressiecoëfficiënt vormen het vaste gedeelte van het model. De drie variantiecomponenten vormen het random gedeelte. Voor het model kan opnieuw de verhouding tussen de variantiecomponenten worden berekend. Deze maat wordt nu de *residuele intra-klasse correlatiecoëfficiënt* genoemd. De correlatie is gelijk aan de correlatie tussen de scores van twee micro-units in dezelfde macro-unit, *gegeven variabele X*.

De onafhankelijke variabele X kan op een hoger niveau differentiële effecten hebben. Vandaar dat er als derde stap moet worden gekeken of er op klas- of schoolniveau *random effects* zijn. Een model met een random intercept en een random helling van de predictorvariabele ziet er als volgt uit:

$$(2.3) Y_{ijk} = \beta_{0jk} + \beta_{1jk} x_{ijk} + R_{ijk}$$

met de volgende uitwerkingen:

$$(2.4) \beta_{0jk} = \gamma_{00k} + U_{0jk}$$

$$(2.5) \beta_{1jk} = \gamma_{10k} + U_{1jk}$$

$$(2.6) \gamma_{00k} = \delta_{000} + V_{00k}$$

$$(2.7) \gamma_{10k} = \delta_{100} + V_{10k}$$

Het *random intercept* op klasniveau (2.4) is opgebouwd uit een schaalgemiddelde en een klasspecifieke afwijking. Op schoolniveau bestaat het random intercept (2.6) eveneens uit een schaalgemiddelde en een afwijking. De *random helling* (2.5) duidt erop dat er voor bepaalde groepen leerlingen, bijvoorbeeld leerlingen met een hoge intelligentie in vergelijking met leerlingen met een lage intelligentie, differentiële effecten zijn. Naarmate \bar{x}_{ijk} groter is, zijn er in een klas grotere verschillen tussen leerlingen naar gelang hun intelligentie. Klassen kunnen dus verschillen in het gemiddelde prestatieniveau en ook naar de relatie van intelligentie met prestaties. Op schoolniveau houdt de *random helling* (2.7) in dat de relatie van intelligentie met prestaties verschilt van school tot school. Beide random hellingen hebben ook een component voor afwijking, zoals in de formules te zien is. Substitutie van de vier formules in 2.3 leidt tot een vergelijking met een intercept, een vast effect en vijf random effecten:

$$(2.8) Y_{ijk} = \delta_{000} + \delta_{100} x_{ijk} + V_{00k} + V_{10k} x_{ijk} + U_{0jk} + U_{1jk} x_{ijk} + R_{ijk}$$

Als vierde stap kunnen *effects* van variabelen op *hoger niveau* worden toegevoegd, bijvoorbeeld een vast effect van de grootte van een klas of het aantal allochtone leerlingen op een school. Aan de indices kan worden gezien op welk niveau de variabelen variëren.

$$(2.9) Y_{ijk} = \delta_{000} + \delta_{100} x_{ijk} + \gamma_{010} x_{0jk} + \delta_{001} x_{00k} + V_{00k} + V_{10k} x_{ijk} + U_{0jk} + U_{1jk} x_{ijk} + R_{ijk}$$

Het model bestaat nu uit een vast gedeelte met een intercept en drie vaste effecten en een random gedeelte met vijf effecten. Dit model is vooral ter illustratie. In de komende hoofdstukken zal blijken dat de gehanteerde multiniveau modellen meer

vaste dan random effecten bevatten. De random effecten bevinden zich in die modellen met name op het klasniveau en niet op het schoolniveau.

Dit model kan tot slot nog worden uitgebreid door *multiniveau interacties* (cross-level interactions) op te nemen, bijvoorbeeld tussen een variabele op klasniveau (x_{0jk}) en een variabele op individueel niveau (x_{ijk}):

$$(2.10) \quad Y_{ijk} = \delta_{000} + \delta_{100}x_{ijk} + \delta_{010}x_{0jk} + \delta_{001}x_{00k} + \delta_{110}x_{0jk}x_{ijk} \\ + V_{00k} + V_{10k}x_{ijk} + U_{0jk} + U_{1jk}x_{ijk} + R_{ijk}$$

Het vaste gedeelte van het model bestaat uit een intercept, drie hoofdeffecten en een interactie-effect. Het random gedeelte bestaat wederom uit vijf effecten.

Multivariate analyse

Een multivariaat model is een model met twee of meer afhankelijke variabelen per onderzoekselement. In dit onderzoek gebruik ik twee criteriumvariabelen, te weten de toetsscores Nederlands en wiskunde. Deze afhankelijke variabelen bevinden zich op niveau nul van het multiniveau model, het pseudo-niveau.

Hoewel het eenvoudiger is om voor elke afhankelijke variabele een aparte analyse uit te voeren, is er een aantal voordelen van een multivariate aanpak (Duncan, Jones & Moon, 1996; Goldstein, 1995²; Snijders & Bosker, 1999). Een belangrijk pluspunt van multivariate modellen is dat kan worden nagegaan of variabelen een *vergelijkbaar of verschillend effect* hebben op de afhankelijke variabelen. Er kunnen significantietoetsen worden uitgevoerd om na te gaan of de effecten verschillen. In de volgende hoofdstukken zal dit voordeel veelvuldig worden benut. In de tweede plaats kan de *correlatie* tussen de afhankelijke variabelen *op verschillende niveaus* worden bepaald. Behalen dezelfde leerlingen goede prestaties op Nederlands en wiskunde? Zijn klassen waar de prestaties gemiddeld op het ene vak hoog zijn ook de klassen waar de prestaties op het andere vak hoog zijn? En hoe ziet deze relatie eruit op schoolniveau? Deze correlaties kunnen niet worden uitgerekend wanneer er aparte analyses voor Nederlands en wiskunde worden uitgevoerd. Tot slot hoeven de respondenten niet op alle afhankelijke variabelen een score te hebben. *Ontbrekende waarden* spelen dus minder een probleem bij multivariate analyse. Voor dit onderzoek geldt dus dat leerlingen in de analyse kunnen worden meegenomen als ze een score hebben op ten minste één van de twee toetsen. Hierbij geldt dat leerlingen die maar op één van de toetsen een score hebben alleen maar meedoen in de analyse van die toets. Door de covariantie tussen beide toetsen wordt de analyse van de andere toets iets nauwkeuriger in de multivariate aanpak.

Om de multivariate analyse uit te kunnen voeren, moeten de afhankelijke variabelen genest worden binnen leerlingen. Per leerling moet dan voor elke criteriumvariabele een record worden gemaakt. In een analyse met één afhankelijke variabele ziet een datamatrix er als volgt uit:

Tabel 2.1: Datamatrix voor multiniveau analyse

NUMMER	KLAS	SCHOOL	NEDERLANDS	WISKUNDE	ETNICITEIT	SES
1	1	1	115	.	0	3
2	1	1	100	85	1	2
.						
.						
6896	452	149	112	130	0	4

In dit onderzoek doen 6.896 leerlingen mee. Deze leerlingen hebben een score op ten minste een van de beide prestatietoetsen. Leerling 1 heeft een ontbrekende waarde op wiskunde. Voor multivariate analyse moet deze matrix worden omgezet naar de volgende structuur:

Tabel 2.2: Datamatrix voor multivariate multiniveau analyse

NUMMER	KLAS	SCHOOL	PRESTATIE	INDICATOR	ETNICITEIT	SES
1	1	1	115	1	0	3
2	1	1	100	1	1	2
2	1	1	85	2	1	2
.						
.						
6896	452	149	112	1	0	4
6896	452	149	130	2	0	4

In deze structuur zijn de scores op Nederlands en wiskunde samengevoegd. De scores kunnen van elkaar worden onderscheiden door een indicator die 1 is voor Nederlands en 2 voor wiskunde. Het pseudo-niveau wordt dus geschapen door gebruik te maken van een dummy. Voor de multivariate analyse zijn er maximaal 13.792 onderzoekselementen, dat wil zeggen twee voor elk van 6.896 leerlingen. Aangezien sommige leerlingen niet beide toetsen hebben gemaakt, zal het aantal onderzoekselementen wat lager uitvallen.

In een multivariate multiniveau analyse met twee afhankelijke variabelen genest binnen leerlingen binnen klassen binnen scholen, is Y_{hijk} de afhankelijke va-

riabele. Het is de meting van afhankelijke variabele h , voor persoon i binnen klas j en school k . Wanneer het random intercept model (formule 2.2) wordt uitgebreid voor de multivariate aanpak, dan is het resultaat:

$$(2.11) Y_{hijk} = \beta_{h000} + \beta_{h100} x_{1hijk} + V_{h00k} + U_{h0jk} + R_{hijk}$$

Vergelijking 2.11 bestaat dus voor elk van de afhankelijke variabelen uit een intercept, een regressiecoëfficiënt en drie residuele variantiecomponenten. Er is een random gedeelte op het leerling-, klas- en schoolniveau, maar niet op het pseudo-niveau. Stel dat de onafhankelijke variabele X in dit multivariate model het aantal dagen is dat een leerling huiswerk maakt, dan worden er voor deze predictor twee parameters geschat: één voor het effect op Nederlands en één voor het effect op wiskunde. Door een chi-kwadraattoets uit te voeren kan worden nagegaan of deze effecten significant verschillen (Goldstein, 1995²). Wanneer dat niet het geval is, kan worden volstaan met één parameter voor het effect van het maken van huiswerk op beide afhankelijke variabelen.

Op de drie niveaus met een residuele variantiecomponent bestaat het random gedeelte uit de variantie van elk van de afhankelijke variabelen plus de covariantie tussen de afhankelijke variabelen. Met deze drie schattingen kan op elk niveau de correlatie tussen de criteriumvariabelen worden berekend. Op het schoolniveau is dit bijvoorbeeld:

$$(2.12) \rho(U_{100k}, U_{200k}) = \frac{\text{cov}(U_{100k}, U_{200k})}{sd(U_{100k})sd(U_{200k})}$$

Gegeven de opgenomen verklarende variabelen, geeft deze correlatie aan of scholen die gemiddeld genomen goed scoren op Nederlands ook de scholen zijn die gemiddeld genomen goed scoren op wiskunde.

Significantietoetsen

Om na te gaan of de *vaste effecten* significant van nul afwijken wordt de parameterwaarde gedeeld door de bijbehorende standaardfout. Dit quotiënt is een t-waarde. Hoewel er duidelijke hypothesen zijn over de richting van effecten, wordt in dit onderzoek voor het gemak tweezijdig getoetst. Op die manier vormt het voor de weergave van de significantie van effecten geen probleem dat een aantal effecten een richting heeft die tegengesteld is aan de verwachte richting. De toetsing is hierdoor aan de voorzichtige kant voor de meeste effecten. Voor een eenzijdige toetsing is een effect dat kleiner is dan $-1,645$ of groter is dan $1,645$ significant bij een alfa van 0,05 en meer dan duizend vrijheidsgraden; bij een tweezijdige toetsing moet de t-waarde absoluut gezien groter dan 1,96 zijn.

Voor de *random effecten* kan er geen gebruik worden gemaakt van t-toetsen. Hier kan echter het principe van de deviance-toets worden gehanteerd. Om twee modellen te vergelijken, moet van beide modellen de deviance worden uitgerekend. De deviance van een model is gelijk aan -2 maal de natuurlijke logaritme van de gezamenlijke kansdichtheid van de waarnemingen, waarbij de via de maximum-likelihood-methode geschatte parameters worden ingevuld. Om na te gaan of de opname van één of meer variabelen tot een verbetering van een model leidt, kan worden gekeken naar het *verschil in deviance* tussen de modellen. Dit verschil is chi-kwadraat verdeeld.

Verklaarde variantie

De variantie in Y kan worden gesplitst in een deel dat door de vaste effecten wordt verklaard en een deel dat onverklaard blijft. De verklaarde variantie kan worden berekend door de proportionele reductie in de onverklaarde variantie te berekenen (Snijders & Bosker, 1999). Voor niveau één tot en met drie is dit in de multivariate aanpak:

$$(2.13) R_{h1}^2 = 1 - \frac{\text{var}(Y_{hijk} - \sum_g \delta_g X_{ghijk})}{\text{var}(Y_{hijk})}$$

$$(2.14) R_{h2}^2 = 1 - \frac{\text{var}(\bar{Y}_{h.jk} - \sum_g \delta_g \bar{X}_{gh.jk})}{\text{var}(\bar{Y}_{h.jk})}$$

$$(2.15) R_{h3}^2 = 1 - \frac{\text{var}(\bar{Y}_{h..k} - \sum_g \delta_g \bar{X}_{gh..k})}{\text{var}(\bar{Y}_{h..k})}$$

Voor niveau één is de verklaarde variantie de proportionele reductie in de som van Φ_v^2 , Φ_u^2 en Φ_r^2 door de opname van één of meer onafhankelijke variabelen. Voor niveau twee is het de proportionele reductie in de som van Φ_v^2 , Φ_u^2 en Φ_r^2/n . De variantie op het individuele niveau wordt gedeeld door het aantal onderzoekselementen binnen niveau twee. In dit onderzoek verschilt het aantal leerlingen van klas tot klas. Bij dergelijke ongebalanceerde data is het verstandig om een representatieve waarde voor het aantal leerlingen te nemen, bijvoorbeeld $n=20$. De verklaarde variantie voor niveau drie is de proportionele reductie in de som van Φ_v^2 , Φ_u^2/j en Φ_r^2/nj , waarbij j het totaal aantal klassen en nj het totaal aantal leerlingen per school is. Aangezien het aantal klassen van school tot school verschilt, moet opnieuw een representatieve waarde worden gekozen, bijvoorbeeld $j=3$ (Snijders & Bosker, 1999).

2.3 Scholen op rapport

In de vorige paragraaf heb ik uitgelegd wat multiniveau analyse in grote lijn inhoudt. Deze uitleg vormt de technische achtergrond voor dit onderzoek. In de hoofdstukken 3, 4 en 5 gebruik ik multivariate multiniveau analyse om de verschillende invloeden op de prestaties en vorderingen van leerlingen te verklaren. In deze paragraaf zal ik uiteenzetten dat in het Trouw-onderzoek multiniveau analyse had moet worden gebruikt, omdat er sprake is van een hiërarchie van gegevens. Hierbij gebruik ik een hiërarchische structuur met een individueel en een schoolniveau. In de volgende hoofdstukken neem ik ook het klasniveau op. De probleemstelling is: *In hoeverre verschillen de schattingen van de netto prestaties van scholen in een analyse met geaggregeerde gegevens en een multiniveau analyse?*ⁱⁱⁱ

Rapportcijfer

Het rapportcijfer uit het Trouw-onderzoek is een cijfer voor het totale rendement van een school, of eventueel van een schooltype binnen een school (Dronkers, 1997b). Dit rendement is een gewogen optelsom van examencijfers, het slagingspercentage, het uitvalpercentage (twee typen) en het percentage zittenblijvers. Voor vwo-, havo-, mavo- en vbo-afdelingen worden respectievelijk vijftien, dertien, tien en negen kenmerken gehanteerd.

Per schooltype is een rapportcijfer samengesteld, waarbij geldt dat het rapportcijfer hoger is naarmate de eindexamencijfers hoger zijn, het slagingspercentage hoger is en de percentages zittenblijvers en uitvallers lager zijn. Het rapportcijfer is daarbij gecorrigeerd voor het percentage allochtone leerlingen op een school. Indien de rendementcijfers gelijk blijven, neemt voor bijvoorbeeld het vbo het rapportcijfer met één punt toe bij 25 procent meer allochtone leerlingen. Bij 12 procent meer allochtone leerlingen neemt het vwo-rapportcijfer met dezelfde hoeveelheid toe.

Voor de berekening van het rapportcijfer is onder meer factoranalyse en regressieanalyse gebruikt. In het onderstaande kom ik op deze analysemethoden terug. Uit de uiteindelijke analyses komt per school een score. Bij die score worden zes punten opgeteld, zodat de gemiddelde school een zes krijgt. Scholen met een vijf of lager behoren tot het kwart laagst scorende scholen. Scholen met een zeven of meer maken deel uit van het kwart hoogst scorende scholen. Als laatste stap zijn de rapportcijfers afgerond op hele cijfers.^{iv} Wanneer er binnen een bepaald schooltype minder dan vijftig eindexamenkandidaten zijn, heeft die school geen rapportcijfer gekregen.

Om de onderzoeksvraag naar de betrouwbaarheid van het rapportcijfer van het Trouw-onderzoek te beantwoorden volg ik een tweevoudige strategie. In eerste instantie volg ik de door Trouw gehanteerde analysemethode, een combinatie van factor- en regressieanalyse, en ga ik na wat de effecten zijn van verschillende manieren om te controleren voor de kenmerken van de leerlingenpopulatie. In tweede instantie wordt gebruik gemaakt van multiniveau analyse, eerst met de variabelen waarop het rapportcijfer van Trouw is gebaseerd en vervolgens met een alternatieve correctie voor instroomverschillen tussen scholen. De analyse omvat derhalve vier stappen:

1A: methode Trouw - correctiefactor Trouw

1B: methode Trouw - alternatieve correcties

2A: alternatieve methode - correctiefactor Trouw

2B: alternatieve methode - alternatieve correcties

Cohortonderzoek VOCL'89

In deze replicatie maak ik geen gebruik van de onderwijsindicatoren die de inspectie van alle scholen in Nederland verzamelt en waarop het onderzoek van Trouw is gebaseerd. Een belangrijk nadeel van die gegevens is de beperkte beschikbaarheid van informatie over de achtergrondkenmerken van leerlingen. Bij het berekenen van een rapportcijfer kon Trouw namelijk alleen voor het percentage allochtone leerlingen corrigeren. Een recente dataset met informatie over schoolloopbanen in het voortgezet onderwijs waarin wel uitgebreide leerlinggegevens beschikbaar zijn, is het cohortonderzoek VOCL'89. Aangezien ik niet zozeer geïnteresseerd ben in de rapportcijfers van te identificeren scholen, maar in de vraag in hoeverre de in het Trouw-onderzoek gekozen analysemethode en gebruikte correctie effect hebben op de beoordeling van scholen, is de anonimiteit van de scholen in het VOCL-bestand geen bezwaar. De navolgende analyses zijn derhalve op dit cohort gebaseerd. Hierbij gebruik ik gegevens van 19.260 leerlingen van 381 scholen.^v

Voor dit cohortonderzoek hebben het Centraal Bureau voor de Statistiek (CBS), het Gronings Instituut voor Onderwijsonderzoek (GION) en het Nijmeegs Instituut voor Toegepast Sociaal-wetenschappelijk onderzoek (ITS) gegevens verzameld over scholen, docenten, leerlingen en ouders. Deze gegevens worden in eerste instantie gebruikt voor de evaluatie van de basisvorming en het onderwijsvoorrangsbeleid, maar kunnen ook voor secundaire analyses worden gebruikt. Driessen en Van der Werf (1992) bespreken de selectie en de representativiteit van de steekproef, waarbij het van belang is dat de steekproef representatief is voor alle scholen voor voortgezet onderwijs in Nederland.

De 19.260 leerlingen zaten in het schooljaar 1989-'90 in de eerste klas van het voortgezet onderwijs. In het schooljaar 1996-'97 volgde 42 procent van deze leerlingen nog onderwijs. Deze leerlingen volgden voornamelijk middelbaar beroeps-, hoger beroeps- of wetenschappelijk onderwijs. Met een diploma heeft 48 procent het onderwijs verlaten. Eentiende van het cohort is zonder diploma uit het onderwijs vertrokken. In Tabel 2.3 vat ik deze gegevens samen.

Tabel 2.3: Positie leerlingen uit VOCL'89 in het schooljaar 1996-'97 (N=19.260)

Leerling:

- volgt momenteel onderwijs	42.2%	8.136
- heeft onderwijs met diploma verlaten	48.0%	9.254
- heeft onderwijs zonder diploma verlaten	9.7%	1.870

Prestatiematen

Er worden vier maten voor prestaties gehanteerd, te weten of een leerling onvertraagd het examen heeft gehaald, het aantal keren dat een leerling is blijven zitten, of een leerling al dan niet voortijdig de school heeft verlaten en de positie van een leerling op een zogeheten leerjarenladder. Deze maten bereken ik per schooltype, waarbij de richtlijn is dat er minimaal tien leerlingen binnen één schooltype vallen. De scores op deze maten zijn op individueel niveau in Tabel 2.4 weergegeven. In Bijlage 2A zijn deze scores op schoolniveau afgebeeld.

Om te bepalen of een leerling al dan niet *onvertraagd geslaagd* is, ben ik voor vwo, havo, mavo en vbo nagegaan of een leerling op dat niveau examen heeft afgelegd.^{vi} Leerlingen zijn onvertraagd geslaagd, wanneer zij het voorlaatste en laatste leerjaar in één keer hebben doorlopen. Leerlingen zijn vertraagd geslaagd, wanneer zij in het voorlaatste leerjaar zijn blijven zitten of wanneer zij voor hun examen zijn gezakt. Een leerling die na zijn mavo-examen verder gaat op de havo en daar ook examen doet, krijgt voor beide niveaus een score op deze prestatie maat. Het percentage onvertraagd geslaagden is het hoogste bij mavo-leerlingen, 86 procent, en het laagste bij havo-leerlingen, 67 procent.

Tabel 2.4: Prestaties op individueel niveau

VBO	GEM.	SD	MIN.	MAX.	N
onvertraagd geslaagd	0,75	0,43	0	1	6.357
zittenblijvers leerjaar 2	0,10	0,30	0	2	7.191
zittenblijvers leerjaar 3	0,13	0,35	0	3	7.191
uitvallers leerjaar 2	0,03	0,17	0	1	7.191

uitvallers leerjaar 3	0,08	0,26	0	1	7.191
positie leerjarenladder	6,3	1,9	2	11	7.191
Mavo					
onvertraagd geslaagd	0,86	0,34	0	1	6.348
zittenblijvers leerjaar 2	0,11	0,32	0	2	6.131
zittenblijvers leerjaar 3	0,15	0,36	0	2	6.131
uitvallers leerjaar 2	0,02	0,16	0	1	6.131
uitvallers leerjaar 3	0,03	0,18	0	1	6.131
positie leerjarenladder	7,9	1,5	4	11	6.131
Havo					
onvertraagd geslaagd	0,67	0,47	0	1	3.237
zittenblijvers leerjaar 2	0,08	0,26	0	2	3.264
zittenblijvers leerjaar 3	0,10	0,30	0	2	3.264
zittenblijvers leerjaar 4	0,49	0,59	0	2	3.264
uitvallers leerjaar 2	0,01	0,09	0	1	3.264
uitvallers leerjaar 3	0,02	0,12	0	1	3.264
uitvallers leerjaar 4	0,05	0,22	0	1	3.264
positie leerjarenladder	9,0	1,2	5	12	3.264
Vwo					
onvertraagd geslaagd	0,82	0,39	0	1	2.519
zittenblijvers leerjaar 2	0,03	0,16	0	1	2.674
zittenblijvers leerjaar 3	0,03	0,17	0	2	2.674
zittenblijvers leerjaar 4	0,08	0,27	0	1	2.674
zittenblijvers leerjaar 5	0,24	0,44	0	2	2.674
uitvallers leerjaar 2	0,00	0,06	0	1	2.674
uitvallers leerjaar 3	0,01	0,10	0	1	2.674
uitvallers leerjaar 4	0,01	0,12	0	1	2.674
uitvallers leerjaar 5	0,04	0,19	0	1	2.674
positie leerjarenladder	10,4	1,1	6	12	2.674

Voor elke leerling is het aantal keren *zittenblijven* berekend door na te gaan hoe vaak hij een bepaald leerjaar meer dan één keer heeft doorlopen. In navolging van het Trouw-onderzoek telt bij de bepaling van deze prestatie maat doubleren in het eerste of laatste leerjaar niet mee. Een overgang van bijvoorbeeld mavo-4 naar havo-4 of havo-5 naar vwo-5 beschouw ik wel als zittenblijven. Met een voorbeeld licht ik deze keuze toe. In het eerste leerjaar zitten twee leerlingen in een havo/vwo-brugklas. De ene leerling krijgt een havo-advies. Deze leerling haalt in vijf jaar tijd zijn diploma en stroomt door naar vwo-5. De andere leerling gaat na de brugklas naar de vwo-afdeling en blijft zitten in het vierde leerjaar. Het

jaar daarop gaat zij wel over naar vwo-5. In hun zesde jaar van voortgezet onderwijs belanden beide leerlingen dus weer bij elkaar in de klas. Vandaar dat ik hen een zelfde score op deze prestatie maat toeken, namelijk één keer blijven zitten. Van alle leerlingen uit VOCL'89 heeft 62 procent geen enkele klas overgedaan, 32 procent is een keer blijven zitten, 6 procent heeft tweemaal gedoubleerd en een gering aantal leerlingen (76 van de 19.260) heeft drie maal een leerjaar overgedaan. In Tabel 2.4 valt te zien dat het percentage zittenblijvers hoog is in vwo-5 (24 procent) en vooral in havo-4 (49 procent).

Als derde prestatie maat heb ik berekend of een leerling al dan niet tot de *uitvallers* van het voortgezet onderwijs behoort. In navolging van het Trouwonderzoek beschouw ik leerlingen als uitvallers, wanneer ze de school na het eerste en voor het laatste leerjaar hebben verlaten. Van alle leerlingen uit VOCL'89 heeft 10 procent de school voortijdig verlaten. Tot deze groep behoren ook leerlingen die bijvoorbeeld doorstromen van havo-4 naar mbo-1. Voor de meeste uitvallers geldt dat ze geen diploma voor voortgezet onderwijs hebben, maar een klein deel (152 van de 2.022) valt gediplomeerd uit. Dit gaat bijvoorbeeld op voor leerlingen met een mavo-diploma die in havo-4 uitvallen. Deze leerlingen gelden niet als mavo- maar wel als havo-uitvaller. De uitval is hoog in vbo-3 (8 procent), havo-4 (5 procent) en vwo-5 (4 procent), zie Tabel 2.4.

Als vierde prestatie maat heb ik de *leerjarenladder* (Bosker, Van der Velden en Hofman, 1985) gebruikt, zie Tabel 2.5. Met deze leerjarenladder kan ik nagaan welk niveau leerlingen sinds het schooljaar 1989-'90 hebben bereikt. De leerjarenladder drukt de afstand in leerjaren uit tot de top van het onderwijssysteem. Het maximale niveau voor deze groep blijkt twaalf te zijn, zie Tabel 2.4. Dit niveau kan worden bereikt door leerlingen die zonder doubleren het vwo hebben doorlopen en die inmiddels in het tweede leerjaar van het wetenschappelijk onderwijs zitten.^{vii} Bij de opstelling van de leerjarenladder hebben Bosker, Van der Velden en Hofman rekening gehouden met zittenblijven en doorstroom. Een leerling die blijft zitten verliest een jaar op de overigen. Bij doorstroming van bijvoorbeeld mavo-4 naar havo-4 doet een leerling een jaar langer over het bereiken van dat niveau dan leerlingen die de havo regulier doorlopen. Bij de bepaling van de positie op de leerjarenladder valt jammer genoeg geen onderscheid te maken tussen de verschillende niveaus binnen de mavo en het vbo. Leerlingen die de mavo afsluiten met alle vakken op D-niveau krijgen dezelfde positie op de leerjarenladder als leerlingen die vakken op C-niveau afleggen. In Tabel 2.4 valt te zien dat naarmate het schooltype hoger is, de positie op de leerjarenladder gemiddeld genomen ook hoger is. Vwo-leerlingen scoren gemiddeld 10,4 en vbo-leerlingen gemiddeld 6,3.

*Tabel 2.5: Leerjarenladder en bijbehorende onderwijsposities
(gebaseerd op Bosker, Van der Velden & Hofman, 1985)*

WAARDE	ONDERWIJSPOSITIE				
14	wo-4				
13	wo-3				
12	wo-2	hbo-4			
11	wo-1	hbo-3			
10	vwo-6	hbo-2	mbo-4		
9	vwo-5	hbo-1	mbo-3		
8	vwo-4	havo-5	mbo-2		
7	vwo-3	havo-4	mbo-1	kmbo-2	
6	vwo-2	havo-3	mavo-4	kmbo-1	
5	vwo-1	havo-2	mavo-3	vbo-4	
4		havo-1	mavo-2	vbo-3	ivbo-4
3			mavo-1	vbo-2	ivbo-3
2				vbo-1	ivbo-2
1					ivbo-1

Netto prestaties

Voor de beoordeling van scholen mogen verschillen in leerlingenpopulatie niet uitmaken. Vandaar dat de bruto rapportcijfers moeten worden gecorrigeerd voor enkele kenmerken van de leerlingenpopulatie. In deze replicatie doe ik dat voor de sociaal-economische status van het gezin, de etniciteit en het beginniveau van een leerling. Voor de sociaal-economische status hanteer ik drie kenmerken, te weten de opleiding van de vader, de opleiding van de moeder en het beroep van de kostwinner. Bij de *opleiding van ouders* onderscheid ik zes eindniveaus, te weten: (1) minder dan basisonderwijs; (2) basisonderwijs; (3) lagere vormen van voortgezet onderwijs (lager beroepsonderwijs, mavo of de onderbouw van havo en vwo); (4) hogere vormen van voortgezet onderwijs en middelbaar beroepsonderwijs; (5) universitair en hoger beroepsonderwijs; (6) postdoctorale universitaire opleidingen (inclusief doctoraal/ingenieursexamen oude stijl). Bij het *beroepsniveau van de kostwinner* maak ik gebruik van een indeling in vijf categorieën: (1) mensen zonder betaald werk; (2) arbeiders; (3) zelfstandigen, landbouwers en lagere employés; (4) middelbare employés; (5) hogere employés en vrije beroepen. Om de *etniciteit* van leerlingen te bepalen heb ik gekeken naar het geboorteland van leerlingen en hun ouders. Hierbij onderscheid ik de volgende groepen: (1) Nederlands; (2) Westers en overig (bijvoorbeeld Duits, Amerikaans,

Pools, Chinees); (3) Turks; (4) Marokkaans; (5) Postkoloniale immigranten (Antilliaans, Arubaans, Surinaams en Moluks). Wanneer leerlingen tot een van de laatste drie groepen behoren, noem ik hen allochtoon. Hierbij hanteer ik als regel dat leerlingen tot een van die groepen behoren als zij zelf of één van hun ouders in Turkije, Marokko of een (voormalige) kolonie is geboren. Dit betekent dus dat ik ook kinderen uit gemengde gezinnen (bijvoorbeeld een Turkse man en een Nederlandse vrouw) als allochtoon beschouw.^{viii} Om de *eerdere prestaties* van leerlingen te bepalen, gebruik ik hun scores op de toetsen rekenen, taal en informatieverwerking die zijn afgenomen in het eerste leerjaar. De maximale score bedraagt zestig punten. Driessen en Van der Werf (1992) beschrijven deze toetsen. In Bijlagen 2B en 2C geef ik enkele beschrijvende gegevens over de kenmerken van de leerlingenpopulatie.

De onderlinge samenhang tussen de kenmerken van de leerlingenpopulatie is op schoolniveau hoog, zie Bijlagen 2D en 2E. De correlaties tussen het gemiddelde beroepsniveau van de kostwinner en het opleidingsniveau van beide ouders is 0,90 of hoger. Met de gemiddelde score op de begintoetsen correleren deze kenmerken van de sociaal-economische status ongeveer 0,80. De kenmerken hebben dus een groot deel van de variantie gemeenschappelijk. Er kan hier dan ook van *multicollineariteit* worden gesproken. Het gevolg van deze sterke intercorrelaties is dat de unieke bijdrage van de afzonderlijke kenmerken gering is. Wanneer ik één van de kenmerken van de leerlingenpopulatie in een regressieanalyse opneem en vervolgens een andere toevoeg, dan zal het extra kenmerk nauwelijks extra variantie kunnen binden. De samenhang tussen het percentage allochtone leerlingen en de overige vier achtergrondkenmerken is minder sterk. De correlaties variëren van –0,34 tot –0,53.

Analyse: stappen 1A en 1B

Als eerste stap repliceer ik het rapportcijfer van Trouw door zowel de methode als de correctie voor het percentage allochtone leerlingen over te nemen (*stap 1A*). Vervolgens pas ik ook een correctie voor andere kenmerken van de leerlingenpopulatie toe (*stap 1B*).

Analoog aan Trouw heb ik met factoranalyse per schooltype een score berekend die staat voor de prestaties van de school. Het doel van de factoranalyse is enerzijds om het latente begrip schoolprestaties te meten door de scores van school op verschillende metingen (percentage onvertraagd geslaagden, percentage zittenblijvers, percentage uitvallers en gemiddelde score op de leerjarenladder) en anderzijds om die verschillende metingen bondig te beschrijven in één maat. Uit de vier factoranalyses blijkt dat de eerste factor steeds het beste het succes van een school weergeeft. Op deze factor laden de percentages zittenblijvers en uitvallers

positief. Het percentage onvertraagd geslaagden en de gemiddelde score op de leerjarenladder hebben een negatief gewicht op de factor.^{ix} In Bijlage 2F geef ik de ladingen van deze kenmerken op deze factor weer. De eerste factor bindt respectievelijk 35,6; 41,5; 27,4 en 27,6 procent aan variantie voor vbo, mavo, havo en vwo.^x De scores op de factor, die ik voortaan als *succesmaat* aanduid, voor de scholen uit VOCL'89 variëren overigens van een drie tot een acht. Het gemiddelde is een zes en de standaarddeviatie is één. Scholen die extreem hoog of laag scoren komen in deze steekproef niet voor.

Tabel 2.6: Regressieanalyse van de succesmaat op het percentage allochtone leerlingen (1A) en de gemiddelde score op de begintoetsen (1B)

	CONSTANTE	% ALLOCHTONEN	BEGINTOETS	R ²
vbo (N=176)	0,124 (0,091)	-0,014 (0,007) *		0,02
	-1,049 (0,358)	-0,007 (0,007)	0,042 (0,013) **	0,07
mavo (N=208)	0,226 (0,073)	-0,046 (0,007) **		0,17
	-0,391 (0,425)	-0,043 (0,008) **	0,017 (0,012)	0,17
havo (N=74)	0,268 (0,108)	-0,052 (0,012) **		0,20
	-1,113 (0,886)	-0,043 (0,013) **	0,033 (0,021)	0,22
vwo (N=75)	0,236 (0,117)	-0,060 (0,016) **		0,15
	-0,968 (1,054)	-0,050 (0,018) **	0,027 (0,023)	0,16

Met regressieanalyse ben ik per schooltype allereerst nagegaan wat de invloed is van het percentage allochtone leerlingen op de succesmaat. Uit de regressieanalyses komt naar voren dat het percentage allochtone leerlingen steeds significant van invloed is op de schoolprestatie, zie Tabel 2.6. Hierbij blijkt dat deze invloed het zwakste is voor het voorbereidend beroepsonderwijs. Voor alle schooltypen geldt dat naarmate het percentage allochtone leerlingen toeneemt de succesmaat gemiddeld genomen lager is. Het is dan ook eerlijk om de score te corrigeren van scholen met veel leerlingen met een andere etnische afkomst. De verklaarde variantie varieert van 2 procent voor vbo tot 17 procent voor mavo.

In de volgende analyse kijk ik niet alleen naar de invloed van het percentage allochtone leerlingen, maar ook naar die van de eerdere prestaties. De toevoeging van de eerdere prestaties sorteert effect voor het vbo. In de scores op de succesmaat voor vbo-scholen treden veranderingen op. Deze veranderingen zijn maximaal zeventiende punt, wat afgerond neerkomt op mogelijke verschillen van een punt per school. De verklaarde variantie neemt hierbij toe tot 7 procent. Voor de andere schooltypen volstaat het om alleen rekening te houden met het

percentage allochtone leerlingen. De eerdere prestaties van leerlingen hebben geen significante invloed op de schoolprestatie.

Toevoeging van een van de kenmerken van de sociaal-economische status, de opleiding van de vader, leidt alleen voor de havo tot een verbetering van de verklaring van de schoolprestatie, zie Tabel 2.7. De verklaarde variantie neemt toe tot 36 procent en de regressievergelijking verandert behoorlijk. Opmerkelijk genoeg is het effect van de gemiddelde opleiding van de vader op de succesmaat negatief. Gegeven de overige kenmerken, scoren scholen slechter op de succesmaat naarmate ze meer leerlingen hebben met een goed opgeleide vader. Dit verband wordt alleen gevonden, wanneer er wordt gecontroleerd voor de twee andere variabelen.

Tabel 2.7: Regressieanalyse van de succesmaat voor de havo (N=74) op het percentage allochtone leerlingen (stap 1A), de gemiddelde score op de begintoetsen en de gemiddelde opleiding van de vader (stap 1B)

	MODEL 1	MODEL 2
Percentage allochtone leerlingen	-0,052 (0,012) **	-0,070 (0,014) **
Gemiddelde score begintoetsen		0,074 (0,021) **
Gemiddelde opleiding vader		-1,204 (0,290) **
Constante	0,268 (0,108)	1,941 (1,087)
Aangepaste R^2	0,17	0,36

Uit deze uitkomsten blijkt dat de effecten van een meer uitgebreide correctie voor populatieverschillen tussen scholen *voorlopig* meevallen. Door de extra correcties veranderen de scores op de succesmaat voor vbo- en havo-scholen maximaal met een punt, zie de vergelijking tussen de stappen 1A en 1B in Tabel 2.9. De conclusie is dat op schoolniveau de samenhang tussen het percentage allochtone leerlingen, de gemiddelde sociaal-economische status en het gemiddelde prestatieniveau bij intrede zo hoog is dat extra correcties de rapportcijfers nauwelijks veranderen. Kleine verschillen tussen de twee stappen zijn er alleen voor vbo- en havo-scholen.

Prestatie en sturing

Wat kan een school doen om goede examenresultaten te behalen? Er zijn drie mogelijkheden: goed onderwijs geven, goede leerlingen werven of door zitten blijven, afstroom en uitstroom slechte leerlingen wegwerken. Zo beschouwd hebben de resultaten in de klassen vóór de examenklas een dubbel karakter: het

zijn tussenliggende prestatie-maten, maar ook sturingsvariabelen voor de scholen. Een factoranalyse van examenresultaten samen met maten voor uitstroom en zittenblijven geeft de empirische samenhang aan tussen al deze gebruikte variabelen. Maar de mate waarin scholen een streng beleid voor de overgang voeren is op zich niet relevant voor een prestatie-maat. Een scholier die zeven jaar over het vwo doet heeft deze verblijfsduur op school ongeacht of hij een keer is blijven zitten of een keer voor zijn eindexamen is gezakt. In een factoranalyse van examenresultaten met maten voor uitstroom en zittenblijven wordt succes dus gecontamineerd met het door de school gehanteerde sturingsmechanisme. Dit wordt ondersteund door de lage hoeveelheid door de eerste factor verklaarde variantie (zie noot 10). De conclusie moet dan ook zijn dat de in de factoranalyse samengenomen variabelen niet één dimensie vormen. Ze bevatten elementen van zowel echte prestatie als sturingsbeleid.

Het verdient dan ook voorkeur om een succesmaat te gebruiken die niet ten dele een sturingsvariabele is. Voorbeelden van succesmaten die niet deze nadelen hebben zijn het aantal leerlingen dat onvertraagd het eindexamen haalt in het aan het eind van de brugperiode gekozen schooltype en de leerjarenladder. Deze maten behoren dan niet gecombineerd te worden met maten voor uitstroom en zittenblijven. Tegen deze achtergrond heb ik voor de tweede replicatie van het Trouw-rapportcijfer de scores op een leerjarenladder als indicator van de prestaties van scholen gekozen.

Belang van multiniveau analyse

De prestatie-maten van het Trouw-onderzoek zijn gebaseerd op aggregaties van leerlingresultaten. In een onderzoek naar prestaties van leerlingen moet echter rekening worden gehouden met de geneste structuur van de gegevens, in dit geval leerlingen binnen scholen. Aan het geen rekening houden met de hiërarchische structuur van de gegevens is een aantal nadelen verbonden. Het onderscheid tussen de analyse van geaggregeerde data en multiniveau analyse is hier op drie punten relevant, te weten de *nauwkeurigheid* van de resultaten, het onderscheid tussen *binnenschoolse en tussenschoolse regressie-effecten* en het berekenen van *posterior means* (zie ook: Dijkstra, Veenstra, Snijders & Peschar, 1999).^{xi}

Voor variabelen zoals de intelligentie en de culturele en etnische achtergrond van leerlingen kan het effect op individueel niveau, het *binnenschoolse effect*, worden onderscheiden van het effect op schoolniveau, het *tussenschoolse effect*. Bij de analyse van schoolgemiddelden zijn alleen de tussenschoolse effecten te zien. Nu is het mogelijk dat scholen met een 'moeilijke' leerlingenpopulatie, bijvoorbeeld met gemiddeld laag opgeleide ouders of een hoog percentage allochtonen, dat compenseren door goed beleid, zodat het tussenschoolse effect

van hiermee overeenkomende variabelen kleiner is dan het binnenschoolse effect. Wanneer bijvoorbeeld het effect van de opleiding van de ouders op individueel niveau positief is en wanneer scholen met gemiddeld laag opgeleide ouders perfect voor de invloed van dat kenmerk compenseren, dan is bij de analyse van schoolgemiddelden, het effect van opleiding van de ouders nihil. Een kind van laag opgeleide ouders kan dan echter beter op een school zitten waar de gemiddelde opleiding van de ouders laag is! Met andere woorden bij een regressieanalyse op schoolniveau en ervan uitgaande dat sommige scholen goed compenseren voor een 'moeilijke' samenstelling van de leerlingenpopulatie, worden dergelijke scholen onvoldoende voor hun prestatie beloond.

Voor ouders is het dus belangrijk om binnen- en tussenschoolse effecten te kunnen onderscheiden. Dit laat ik in het onderstaande zien voor één variabele X , maar het principe is toepasbaar op een willekeurig aantal variabelen. Wanneer de uitkomstvariabele voor leerling i in school j wordt aangegeven met Y_{ij} , kan ik het effect op Y_{ij} van een variabele X met individuele waarden x_{ij} en schoolgemiddelden $\bar{x}_{.j}$ uitdrukken door de formule:

$$(2.16) Y_{ij} = \beta_{00} + \beta_B (x_{ij} - \bar{x}_{.j}) + \beta_T \bar{x}_{.j} + U_{0j} + R_{ij}$$

Hierin is β_B de binnenschoolse en β_T de tussenschoolse regressiecoëfficiënt. De *onverklaarde* bijdrage van de school, die gecorrigeerd is voor de binnen- en tussenschoolse effecten van X , is weergegeven met U_{0j} . De *onverklaarde* eigen bijdrage van de leerling, voor hetzelfde gecorrigeerd, is R_{ij} . Deze formule komt overeen met de situatie waarin het schoolgemiddelde en de binnenschoolse deviatiescore als variabelen in de analyse worden opgenomen. Het is mogelijk de formule te herschrijven door de oorspronkelijke leerlingvariabele x_{ij} en het schoolgemiddelde $\bar{x}_{.j}$ als variabelen op te nemen:

$$(2.17) Y_{ij} = \beta_{00} + \beta_B x_{ij} + (\beta_T - \beta_B) \bar{x}_{.j} + U_{0j} + R_{ij}$$

Bij een regressie binnen een enkele school geldt op populatieniveau de binnenschoolse regressiecoëfficiënt. Indien de gegevens (zowel X als Y) worden geaggregeerd tot schoolgemiddelden en vervolgens op schoolniveau een regressie wordt uitgevoerd dan geldt de tussenschoolse regressiecoëfficiënt. In dat geval gaat het dus om de grootte van het effect van het schoolgemiddelde X op het schoolgemiddelde Y .

De betekenis van deze coëfficiënten kan als volgt worden toegelicht. Stel bijvoorbeeld dat X de etnische afkomst van een leerling aangeeft: $x_{ij} = 1$ voor allochtone en 0 voor autochtone leerlingen. Dan fungeert de proportie allochtone

leerlingen in school j als het schoolgemiddelde in formule 2.17. Binnen een gegeven school, met een gegeven proportie allochtone leerlingen, is de bijdrage van de etnische achtergrond van een leerling gemiddeld gelijk aan $\bar{\beta}_B$, waarbij de verwachting is dat het effect van etniciteit negatief is (Driessen, 1995; De Jong, 1989). De proportie allochtone leerlingen op een school kan ook een eigen effect hebben op de prestaties van allochtone en autochtone leerlingen. Dergelijke effecten kunnen bijvoorbeeld het gevolg zijn van de invloed van de samenstelling van de leerlingenpopulatie op het groepsproces binnen een klas en de school, bijvoorbeeld door een veranderde tijdsverdeling van docenten en door vergelijkings- en beïnvloedingsprocessen tussen leerlingen. Ook kunnen effecten van het percentage allochtone leerlingen het gevolg zijn van de inspanningen die de school zich getroost om de problemen ontstaan door een 'moeilijke' leerlingenpopulatie het hoofd te bieden. Indien op schoolniveau dit soort effecten zich inderdaad voordoen dan zal de binnenschoolse regressiecoëfficiënt verschillen van de tussenschoolse regressiecoëfficiënt, dus $(\beta_T - \bar{\beta}_B) \neq 0$. Hierbij kan het overigens zo zijn dat deze effecten elkaar opheffen. De 'moeilijke' samenstelling van de leerlingenpopulatie zal een negatief effect hebben op $(\beta_T - \bar{\beta}_B)$, het beleid van de scholen zal, zo is de bedoeling, een positief effect hebben op deze coëfficiënt.

Wat een rapportcijfer voor scholen voor voortgezet onderwijs betreft, zijn er verschillen tussen de standpunten van *ouders* en de *overheid*. Voor ouders is de eigen etnische achtergrond een gegeven. Indien ouders de verwachte schoolprestaties van hun eigen kind willen maximaliseren, zullen zij de scholen willen ordenen op grond van de waarde van:

$$(2.18) (\beta_T - \bar{\beta}_B) \bar{x}_{.j} + U_{0j}$$

Dit is de schoolbijdrage waarbij wordt gecorrigeerd voor de samenstelling van de leerlingenpopulatie door de binnenschoolse en niet door de tussenschoolse regressiecoëfficiënt. Voor ouders wordt dus geen correctie uitgevoerd voor het effect van de leerlingenpopulatie op schoolniveau. Dit is overeenkomstig aan de wensen van ouders. Als hun kind veel leert, omdat de samenstelling van de school gunstig is, dan is dat voor hen mooi meegenomen.

De overheid is niet zo zeer in de prestaties van individuele leerlingen, maar van een groep leerlingen, bijvoorbeeld een school, geïnteresseerd. Voor de overheid is de leerlingenpopulatie van een school dan een gegeven. Indien de overheid de gemiddelde prestaties van de leerlingen op een school wil belonen, zal zij de scholen willen ordenen volgens de waarde van U_{0j} . Dit is immers de

eigen bijdrage van een school, die niet kan worden toegeschreven aan de samenstelling van variabele X . Op deze manier wordt rekening gehouden met zowel de moeilijkheden op schoolniveau door variabele X als de op schoolniveau gemiddelde gevolgen van het beleid waarmee scholen deze moeilijkheden het hoofd bieden. Dat kan ook niet anders, want zonder gegevens over schoolbeleid te gebruiken zijn deze twee invloeden niet van elkaar te isoleren. De scholen met een 'moeilijke' leerlingenpopulatie worden zo dus als groep niet beloond voor hun inspanningen om hier goed mee om te gaan. Alleen wanneer een school het beter of slechter doet dan andere scholen met een zelfde soort leerlingenpopulatie, dan wordt een school beloond of bestraft. Het zou voor de overheid *eerlijker* zijn om wel voor succes op groepsniveau te belonen, maar dit zou hoge eisen stellen aan de te gebruiken data en aan inzichten in de aspecten van het beleid die er toe doen.

Met wat voor analysemethoden kunnen de hier genoemde getallen om de scholen te ordenen worden geproduceerd? Wat de bovenstaande formule vanuit het gezichtspunt van de ouders betreft, is een multiniveau analyse noodzakelijk om afzonderlijk de binnenschoolse regressiecoëfficiënt β_B en de tussenschoolse regressiecoëfficiënt β_T te schatten. Wat het schatten van alleen de voor de overheid relevante waarde U_{0j} betreft zijn er twee mogelijkheden. De gemakkelijkste methode is een regressie van schoolgemiddelden, waaruit U_{0j} als residu kan worden berekend. Dit is analoog aan een onderdeel van de door Trouw uitgevoerde analyse. Een alternatieve methode is weer het schatten van dit residu uit een multiniveau analyse. Dit heeft als voordeel dat met het aantal in de analyse gebruikte leerlingen rekening wordt gehouden, waarbij voor scholen met weinig data over leerlingen een meer voorzichtige schatting wordt gebruikt dan voor scholen met veel data. Bij scholen met een klein aantal leerlingen speelt het toeval namelijk een grotere rol bij de gemiddelde prestaties dan bij scholen met een groot aantal leerlingen, omdat toevalsinvloeden bij de eerste scholen in mindere mate worden weggemiddeld. Dit betekent dat scholen met een relatief klein aantal leerlingen eerder toevallig hoge of lage rapportcijfers zullen krijgen dan scholen met veel leerlingen.

In een multiniveau analyse wordt pas duidelijk hoe groot of klein de nauwkeurigheid van de schoolcijfers is. Wanneer een school bijvoorbeeld een zes als rapportcijfer heeft, dan kan worden bepaald of dat cijfer in het hypothetische geval van een volgende meting, en dus met een andere leerlingengroep, wel eens twee punten hoger of lager kan uitkomen of dat het cijfer waarschijnlijk tussen een 5,5 en een 6,5 zal schommelen. Voor een Engelse hitlijst van scholen laten Goldstein en Spiegelhalter (1996) zien dat de nauwkeurigheid in sommige gevallen tamelijk gering is.^{xii}

Voor zowel ouders als de overheid verdient multiniveau analyse de voorkeur boven een simpele regressieanalyse op geaggregeerde data. Door multiniveau analyse te gebruiken kan voor ouders het noodzakelijke onderscheid tussen binnen- en tussenschoolse regressie-effecten worden gemaakt. Voor de overheid is dit onderscheid van minder gewicht. Een ander voordeel dat wel belangrijk is voor de overheid, betreft de grotere nauwkeurigheid van de standaardfouten en betrouwbaarheidsintervallen in een multiniveau analyse. Verder geldt dat ook scholen met een gering aantal leerlingen een rapportcijfer kunnen krijgen in een multiniveau analyse, omdat dan een meer voorzichtige schatting kan worden gegeven.

Analyse: stappen 2A en 2B

Multiniveau analyse verdient dus de voorkeur boven een op schoolniveau geaggregeerde regressieanalyse. Om na te gaan in hoeverre gebruik van deze in statistisch opzicht meer geëigende methode effect heeft op de succesmaat van scholen, zijn de stappen 2A en 2B gebaseerd op een multiniveau analyse, uitgevoerd met het programma HLM (Bryk, Raudenbush & Congdon, 1993). De score op de leerjarenladder is als succesmaat gekozen. De analyse wordt opnieuw per schooltype uitgevoerd. Om scholen te rangordenen is gebruik gemaakt van de *posterior means*, dat wil zeggen de residuen die voor het intercept op schoolniveau zijn berekend. Met deze empirisch Bayesiaanse schatting kan worden nagegaan welke scholen onverwacht hoog of laag scores op de succesmaat, gegeven de waarden op de onafhankelijke variabelen. Voor ouders is dit de maat die hen informatie geeft over welke scholen voor de gemiddelde leerling succesvol zijn. In stap 2A en 2B gebruik ik deze ‘posterior means’, waarbij deze schattingen zijn gestandaardiseerd.

In stap 2A hanteer ik dezelfde correctie als in het Trouw-onderzoek. Analooq aan de werkwijze van Trouw corrigeer ik alleen voor het percentage allochtone leerlingen van de school. Het verschil is dus enkel de analysemethode. Ik gebruik namelijk multiniveau analyse en als succesmaat de leerjarenladder. De correctie voor het percentage allochtone leerlingen voer ik uit op schoolniveau, terwijl ik op individueel niveau een leeg model schat.

Uit de analyses komt naar voren dat het percentage allochtone leerlingen in drie van de vier gevallen van invloed is op de score op de leerjarenladder, zie Tabel 2.8. Voor havo-leerlingen blijkt er geen invloed te zijn van deze correctiefactor. Voor de overige schooltypen geldt dat naarmate het percentage allochtone leerlingen toeneemt de score op de leerjarenladder gemiddeld genomen lager is. De verklaarde variantie op schoolniveau is 3 procent voor vbo en mavo, en 10 procent voor vwo.

Tabel 2.8: Multiniveau analyse op de leerjarenladder-score voor afzonderlijke schooltypen

STAP 2A (met correctie Trouw)			
	VBO	MAVO	HAVO
Constante	-0,005 (0,016)	-0,015 (0,021)	-0,057 (0,040)
<i>Schoolniveau</i>			
Percentage allochtone leerlingen	-0,003 (0,001) *	-0,007 (0,002) *	0,001 (0,009)
			-0,052 (0,022) *
TUSSENSTAP (uitgebreid model)			
Constante	0,205 (0,064)	-0,417 (0,042)	0,076 (0,044)
<i>Individueel niveau</i>			
Etniciteit (1=allochtoon)	-0,119 (0,044) *	-0,023 (0,039)	-0,189 (0,147)
Opleidingsniveau vader	0,002 (0,009)	0,023 (0,009) *	0,092 (0,015) *
Prestaties bij instroom	0,009 (0,001) *	0,026 (0,002) *	0,039 (0,003) *
<i>Schoolniveau</i>			
Percentage allochtone leerlingen	-0,002 (0,002)	-0,002 (0,002)	0,004 (0,007)
Gemiddelde opleidingsniveau vader	0,033 (0,059)	0,089 (0,056)	-0,148 (0,101)
Gemiddelde prestaties bij instroom	-0,002 (0,004)	-0,001 (0,004)	-0,013 (0,008)
			-0,014 (0,012)
			-0,042 (0,018) *
STAP 2B (spaarzaam model)			
Constante	0,194 (0,024)	-0,347 (0,031)	0,102 (0,042)
<i>Individueel niveau</i>			
Etniciteit (1=allochtoon)	-0,130 (0,042) *		
Opleidingsniveau vader		0,026 (0,009) *	0,083 (0,015) *
Prestaties bij instroom	0,009 (0,001) *	0,026 (0,002) *	0,038 (0,003) *
<i>Schoolniveau</i>			
Percentage allochtone leerlingen			-0,032 (0,013) *

In de volgende analyse voeg ik de variabelen toe waarmee ook andere aspecten van verschillen in leerlingenpopulatie in de succesmaat kunnen worden verdisconteerd. Om een maat te verkrijgen die goed vergelijkbaar is met die uit stap 1B beperk ik me tot de daar meegenomen variabelen. De toevoeging van de etniciteit, het opleidingsniveau van de vader, de prestaties bij instroom (kenmerken op *individueel niveau*), het gemiddelde opleidingsniveau van de vaders, de gemiddelde prestaties bij instroom (kenmerken op *schoolniveau*), blijkt voor alle schooltypen effect te sorteren voor de scores op de leerjarenladder. Het volstaat niet om alleen rekening te houden met het percentage allochtone leerlingen. Op de succesmaat hebben de prestaties bij instroom in alle gevallen een effect. Het opleidingsniveau van de vader blijkt niet van belang voor vbo-leerlingen, maar wel voor leerlingen in het algemeen vormend onderwijs. Voor vbo-leerlingen blijkt het juist van gewicht om te corrigeren voor de etniciteit van leerlingen. De drie kenmerken op individueel niveau hebben in één of meer van de analyses een effect. Van de kenmerken op schoolniveau heeft alleen het percentage allochtone leerlingen een effect en dan alleen voor vwo-leerlingen, zie Tabel 2.8. Net als in de regressieanalyse op geaggregeerde gegevens heeft het gemiddelde opleidingsniveau van de vader een negatief effect op de succesmaat. De regressiecoëfficiënt van het opleidingsniveau op schoolniveau blijkt $-0,148$ te zijn. Op individueel niveau is het effect van de opleiding van de vader $0,092$. Wanneer echter de standaardfouten in beschouwing worden genomen, dan blijkt dat het effect op individueel niveau wel en op schoolniveau niet significant is. Dit wijst erop dat het negatieve effect van het gemiddelde opleidingsniveau van de vader in Tabel 2.7 het gevolg kan zijn van een samenloop van binnen- en tussenschoolse regressie-effecten.

Als tussenstap is voor alle schooltypen een uitgebreid model opgesteld met alle mogelijke effecten op individueel en schoolniveau. Uit deze uitkomsten blijkt dat vooral met kenmerken op individueel niveau en niet zo zeer op schoolniveau rekening moet worden gehouden. Vervolgens zijn in stap 2B alleen de significante effecten meegenomen. De verschillen tussen de in stap 2A en 2B berekende scores op de succesmaat geef ik weer in Tabel 2.9. Het spaarzame model in stap 2B leidt tot aanzienlijke verschillen in de scores van scholen. Voor veel scholen treden verschillen van één of meer punten op. De verschillen zijn het grootst voor mavo- en vwo-scholen. De verklaarde variantie in het spaarzame model varieert op leerlingniveau van 5 procent voor vbo-leerlingen tot 13 procent voor vwo-leerlingen. De verschillen in verklaarde variantie op schoolniveau zijn groot. Op schoolniveau wordt voor havo-leerlingen niets verklaard, maar voor vwo-leerlingen blijkt de helft van de variantie verklaard te worden.

Tabel 2.9: Percentage scholen met verschillende scores op de vier succesmaten

VERSCHIL IN:	VARIABELEN		METHODE	
	1A-1B	2A-2B	1A-2A	1B-2B
vbo				
geen	92	85	53	50
één cijfer	8	15	41	46
twee cijfers	0	0	5	3
meer dan twee cijfers	0	0	1	1
mavo				
geen	100	51	45	34
één cijfers	0	44	44	47
twee cijfers	0	5	6	13
meer dan twee cijfers	0	0	5	6
havo				
geen	99	78	27	22
één cijfer	1	22	52	44
twee cijfer	0	0	21	32
meer dan twee cijfers	0	0	0	2
vwo				
geen	100	55	43	38
één cijfer	0	39	45	36
twee cijfers	0	7	10	26
meer dan twee cijfers	0	0	2	0

Legenda:

1A: methode Trouw - correctiefactor Trouw

1B: methode Trouw - alternatieve correcties

2A: alternatieve methode - correctiefactor Trouw

2B: alternatieve methode - alternatieve correcties

Vergelijking van de vier rapportcijfers

Behalve de verschillen tussen de stappen 2A en 2B, geef ik in Tabel 2.9 de verschillen tussen nog drie stappen weer. De in de tabel weergegeven rapportcijfers zijn de naar z-scores getransformeerde uitkomsten van de analyses uit de stappen 1A tot 2B. Om reden van presentatie heb ik de uiteindelijke verschillen afgerond. Voor de correlaties tussen de scores op de vier succesmaten verwijs ik naar Bijlage 2G. Uit die correlatiematrices valt opnieuw op te maken dat de rapportcijfers uit de stappen 1A en 1B vrijwel gelijk aan elkaar zijn. De correlaties zijn voor alle schooltypen vrijwel gelijk aan 1. De rapportcijfers uit de stappen 2A en 2B hangen sterk met elkaar samen. De correlaties variëren tussen 0,70 en 0,94. Deze samenhang is echter niet perfect. Het maakt voor de rapportcijfers dus wel uit of er met meer of minder factoren wordt rekening gehouden.

De variabelenvergelijking maakt duidelijk dat effecten van verschillen in correctie voor een meer of minder uitgebreide correctie zich vooral voordoen bij de op multiniveau analyse gebaseerde succesmaten (2A-2B). Dergelijke verschillen treden, zoals eerder geconcludeerd, nauwelijks op bij een op schoolniveau geaggregeerde regressieanalyse (1A-1B). Vergelijking van beide methoden (1A-2A, 1B-2B) laat zien dat gebruik van een multiniveau analyse met de leerjarenladder als succesmaat tot een aanzienlijke verschuiving in de scores van scholen leidt.

In Tabel 2.10 vergelijk ik per schooltype de rapportcijfers van enerzijds de op schoolniveau geaggregeerde regressieanalyse met beperkte correctie (stap 1A) en anderzijds de multiniveau analyse met uitgebreide correctie (stap 2B). Voor het vwo geldt bijvoorbeeld dat 39 procent van de scholen in beide stappen hetzelfde cijfer heeft. Deze scholen zijn weergegeven op de diagonaal in Tabel 2.10. De overige vwo-scholen hebben verschillende cijfers in beide stappen, waarbij 19 procent van de scholen een onvoldoende behaalt in de stap die analoog is aan het Trouw-onderzoek en een voldoende scoort in de door mij bepleite stap. Omgekeerd scoort 12 procent van de vwo-scholen in stap 1A voldoende en in stap 2B onvoldoende.

Tabel 2.10: Samenhang in scores tussen de succesmaten uit stap 1A en stap 2B

VBO							MAVO						
		CIJFER STAP 2B							CIJFER STAP 2B				
		4	5	6	7	8			4	5	6	7	8
CIJFER STAP 1A	4	3	3	2	0	0	CIJFER STAP 1A	4	0	3	3	0	2
	5	1	11	10	2	0		5	1	7	9	3	1
	6	2	3	23	7	0		6	0	8	20	6	3
	7	0	1	12	13	1		7	0	5	18	6	1
	8	0	0	1	5	0		8	1	1	1	1	1
HAVO							VWO						
		CIJFER STAP 2B							CIJFER STAP 2B				
		4	5	6	7	8			4	5	6	7	8
CIJFER STAP 1A	4	0	2	2	2	0	CIJFER STAP 1A	4	0	5	2	2	0
	5	0	7	8	13	0		5	2	12	10	5	0
	6	3	21	2	8	0		6	2	5	17	2	0
	7	0	5	7	15	2		7	0	5	12	10	5
	8	0	0	5	0	0		8	0	0	2	0	0

Legenda:
 1A: methode Trouw - correctiefactor Trouw
 2B: alternatieve methode - alternatieve correcties

De keuze van de succesmaat en de analysemethode heeft duidelijke gevolgen voor de scores van scholen. Scholen die met een op schoolniveau geaggregeerde regressieanalyse, de methode van het Trouw-onderzoek, bijvoorbeeld een zes scoren, kunnen met een multiniveau analyse zo maar één of twee punten meer behalen. Deze verschillen zijn het gevolg van een meer adequate controle voor de samenstelling van de leerlingpopulatie. Een voorzichtige interpretatie van rapportcijfers gebaseerd op een berekeningswijze zoals in het dagblad *Trouw* is dan ook geboden.

2.4 Besluit

Het blijkt dat de keuze van de analysemethode grote gevolgen heeft voor de scores van scholen. Scholen die met een op schoolniveau geaggregeerde regressie-analyse, de methode van het Trouw-onderzoek, bijvoorbeeld een zes halen, kunnen met een multiniveau analyse zo maar een punt lager of hoger scoren.

In de verdere ontwikkeling van methoden om de kwaliteit van scholen inzichtelijk te maken, moet worden gepleit voor een multiniveau analyse, omdat alleen met een dergelijke methode adequaat kan worden gecontroleerd voor de samenstelling van de leerlingenpopulatie. Voor zowel ouders als de overheid blijkt multiniveau analyse de voorkeur te verdienen boven een simpele regressieanalyse op geaggregeerde data. Door multiniveau analyse te gebruiken kan voor ouders het noodzakelijke onderscheid tussen binnen- en tussenschoolse regressie-effecten worden gemaakt. Voor de overheid is dit onderscheid van minder gewicht. Een ander voordeel dat wel belangrijk is voor de overheid, betreft de grotere nauwkeurigheid van de resultaten in een multiniveau analyse. Verder geldt dat ook scholen met een gering aantal leerlingen een rapportcijfer kunnen krijgen in een multiniveau analyse, omdat dan een meer voorzichtige schatting kan worden gegeven.

Met multiniveau analyse scholen zijn gemakkelijk te rangordenen door gebruik te maken van zogeheten posterior means. Hierbij verdient het de voorkeur om een leerjarenladder of een soortgelijke maat als indicator van de prestaties van scholen te gebruiken. Een dergelijke maat geeft alleen het succes van scholen weer. In het Trouw-onderzoek zijn variabelen samengenomen die enerzijds het succes en anderzijds de sturingsmechanismen van scholen weergeven. Deze variabelen vormen niet één dimensie. Door de lage hoeveelheid verklaarde variantie van de eerste factor wordt deze conclusie ondersteund.

Wanneer wordt gekozen voor een multiniveau analyse dan blijkt dat de correctie voor alleen het percentage allochtone leerlingen mager is. Het rapportcijfer dient te worden verfijnd met meer gegevens over de kwaliteit van de instroom van scholen. Deze uitgebreide controle heeft eigenlijk vooral zin in een multiniveau analyse.

Bosker e.a. (1998) wijzen ook op het belang van een analyse op individueel en schoolniveau. Ook verwerpt Bosker de factoranalyse, omdat de computer dan bepaalt *in welke mate* factoren, zoals het percentage geslaagden, zittenblijvers, uitvallers en bèta-vakken bijdragen aan de beoordeling van een school (Bosker in Jager, 1998, p. 14). "Dat is alsof je bij een WK schaatsen alle tijden op vier afstanden in de computer stopt. Vervolgens zeg je tegen de deelnemers: 'Wacht even, de computer bepaalt nu hoe zwaar elke afstand voor de uitslag meetelt, en dan weten we zo wie wereldkampioen is.' Als daar dan de 5000 meter als

belangrijkste uitrolt is dat vreemd, want bij het vorige kampioenschap was dat nog de 1500 meter. Aan zo'n koffiemolentechniek kun je een dergelijke beslissing niet toevertrouwen."

In een reactie op Veenstra e.a. (1998) vindt Dronkers (1998a) echter dat de voorkeur voor multiniveau analyse om twee redenen onjuist is. Ten eerste vindt hij dat met de keuze voor die analysemethode te hoog wordt ingezet wat de kwaliteit van data betreft. Aangezien er voorlopig geen gegevens zijn over een voldoende aantal leerlingen van alle scholen voor voortgezet onderwijs, is multiniveau analyse volgens Dronkers (1998a, p. 149) niet toepasbaar. "Als wetenschappers dan toch staan op multilevel analyses, sluiten zij zichzelf uit van het verdere maatschappelijke debat over de meting van 'toegevoegde waarde' van scholen. De wetenschappelijke opvattingen over de favoriete methode hebben dan de bovenhand gekregen, waardoor de maatschappelijke dienstverlening door de wetenschap gehinderd zou worden. In dat geval is het mogelijk betere (een superieur geachte multilevel methode) de vijand van het goede (een min of meer verantwoorde meting van kwaliteit van scholen)." Als aanzet tot een rangordening van scholen mag de maatschappelijke dienstverlening van Dronkers geslaagd worden genoemd. Wetenschappelijk blijft het de vraag of deze manier van rangordenen goed genoeg is. In een vergelijking tussen verschillende methoden blijkt dat in de rapportcijfers van scholen geregeld een verandering van één punt optreedt en dat ook veranderingen van twee punten voorkomen. De resultaten zijn zodanig onnauwkeurig dat het nemen van beslissingen over individuele scholen leidt tot onterechte benadeling (of bevoordeling) van sommige scholen. Bij het opstellen van rangordeningen dient van deze onnauwkeurigheid melding te worden gemaakt.

Ten tweede stelt Dronkers (1998a) dat met multiniveau analyse een andere vraag wordt beantwoord. Inderdaad moet er een verschil worden gemaakt tussen het standpunt van de overheid en dat van ouders. Vanuit het standpunt van de overheid zou het eventueel volstaan om de prestaties van scholen in kaart te brengen met een regressie van schoolgemiddelden, omdat voor de overheid het onderscheid tussen binnen- en tussenschoolse regressie-effecten minder belangrijk is. Om de gegevens van uit het gezichtspunt van ouders te bekijken is het echter noodzakelijk om een multiniveau analyse uit te voeren. Indien ouders de verwachte schoolprestaties van hun eigen kind willen maximaliseren, zullen zij voor de schoolkeuze de scholen willen ordenen op grond van een score van scholen waarbij is gecorrigeerd voor de samenstelling van de leerlingenpopulatie door de binnenschoolse en niet door de tussenschoolse regressiecoëfficiënten. Met het kenmerk etniciteit is dit onderscheid in dit hoofdstuk toegelicht. Voor ouders is het alleen nodig om te controleren voor etniciteit als leerlingkenmerk. In

een op schoolniveau geaggregeerde analyse wordt niet gecontroleerd voor etniciteit op individueel niveau, maar voor het percentage allochtone leerlingen. Dat percentage kan echter ook een eigen effect hebben. Deze effecten kunnen bijvoorbeeld komen door de invloed van de samenstelling van de leerlingenpopulatie op de tijdsverdeling van docenten of de groepsprocessen binnen een school. Ook kunnen effecten het gevolg zijn van de inspanningen die een school verricht om te compenseren voor de 'moeilijke leerlingenpopulatie'. Indien dit soort effecten zich voordoen, dan verschillen de binnen- en tussenschoolse regressie coëfficiënten en worden sommige scholen ten onrechte te veel of te weinig beloond voor hun presteren. Het Trouw-onderzoek bekijkt de prestaties dus niet vanuit het gezichtspunt van ouders, hoewel de redactie dat wel duidelijk voor ogen had.

Eindnoten

ⁱ Voor een multiniveau model worden ook andere termen gebruikt, zoals een random-coëfficiëntmodel, een hiërarchisch lineair model of een variantiecomponentmodel.

ⁱⁱ Als leerlingen in zeven jaar tijd een vwo-diploma behalen, dan doet hun school het volgens het Trouw-onderzoek slechter dan wanneer leerlingen in vier jaar voor hun mavo-examen slagen. Prick (1997) zegt hierover: "De school die streng selecteert, die vindt dat alleen een smalle top in staat is een vwo-diploma te halen, krijgt hoge cijfers; de school die ook minder getalenteerden, wellicht moeizamer, maar toch aan zo'n diploma helpt wordt gediskwalificeerd. Dat is natuurlijk de omgekeerde wereld."

ⁱⁱⁱ De onderstaande tekst is een bewerking van twee artikelen (Dijkstra e.a., 1999; Veenstra e.a., 1998).

^{iv} Dronkers heeft de cijfers naar eigen zeggen afgerond om de indruk van exactheid te voorkomen. Die indruk moet met deze cijfers inderdaad niet worden gewekt. De keuze om dan maar af te ronden is niet gelukkig, omdat kleine verschillen tussen scholen dan groter kunnen lijken dan ze werkelijk zijn.

^v Het Trouw-onderzoek laat zitten blijven of uitvallen in het eerste leerjaar van het voortgezet onderwijs buiten beschouwing. In lijn hiermee neem ik de 262 leerlingen die in het eerste leerjaar zijn uitgevallen niet in de analyses mee.

^{vi} Behalve de gebruikelijke categorieën van vbo, mavo, havo en vwo worden er in VOCL'93 ook enkele andere categorieën gehanteerd, te weten voortgezet speciaal onderwijs, brugklas smal, brugklas breed en middenschool. In het vierde leerjaar worden die categorieën voor 1 procent van de leerlingen gebruikt. De categorie 'brugklas smal' is gelijkgesteld aan havo. De categorieën 'brugklas breed' en 'middenschool' zijn gelijkgesteld aan mavo. De categorie 'voortgezet speciaal onderwijs' is gelijkgesteld aan vbo. Bij het bepalen van de positie op de leerjarenladder worden leerlingen uit het voortgezet speciaal onderwijs overigens als ivbo-leerlingen beschouwd.

^{vii} Dronkers (1999) merkt op dat de leerjarenladder beter 'stilgezet' had kunnen worden wanneer leerlingen hun school voor voortgezet onderwijs verlaten. De vervolgloopbaan van leerlingen is immers het *extern rendement* van scholen.

^{viii} Het effect van etniciteit wordt hierdoor mogelijk enigszins onderschat. Deze operationalisatie is echter gebruikelijk bij de rapportages over VOCL'89 en VOCL'93. Het zou echter goed zijn om de gevolgen van deze keuze na te gaan.

^{ix} Ik heb als tussenstap het teken van de factorscore veranderd, zodat een hoge score inhoudt dat er veel onvertraagd geslaagden, weinig zittenblijvers en weinig uitvallers zijn en de gemiddelde positie op de leerjarenladder hoog is.

^x Het percentage verklaarde variantie is bij de inspectie-gegevens overigens lager dan bij de VOCL-data. De factor bindt respectievelijk 23,5, 25,1, 20,3 en 14,9 procent aan variantie voor vbo, mavo, havo en vwo.

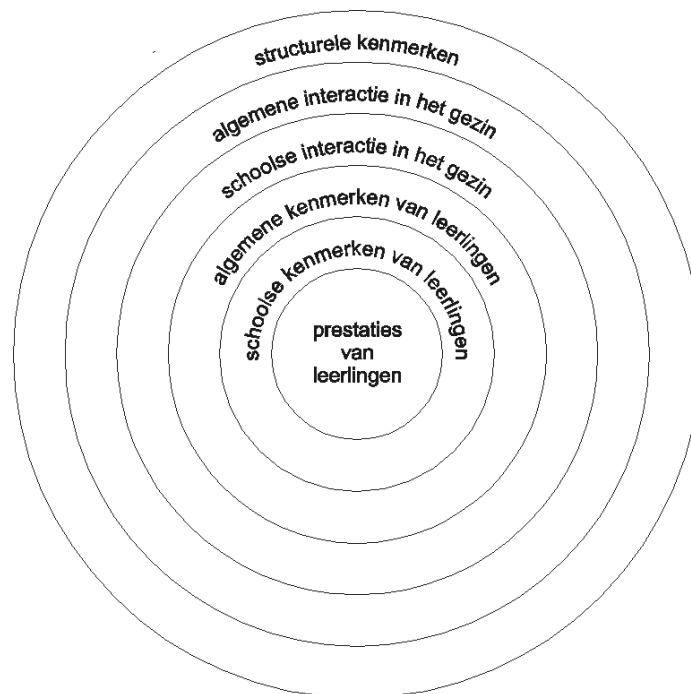
^{xi} De verschillen tussen de analyse van geaggregeerde data en multiniveau analyse zullen vooral aan het licht komen, wanneer de scores van scholen zijn gebaseerd op de gegevens van weinig leerlingen. Snijders en Bosker (1999) geven aan dat de verschillen tussen de methoden gering zijn als er meer dan honderd leerlingen van elke school in het onderzoek zijn betrokken.

^{xii} Dijkstra e.a. (1999) gaan in op de onnauwkeurigheid in de ranglijsten voor Nederlandse scholen voor voortgezet onderwijs. Uitgaande van honderd eindexamenleerlingen binnen een schooltype geldt dat slechts een op de vier scholen significant afwijkt van het gemiddelde. Van deze 25 procent wijkt de helft positief en de andere helft negatief af.

3 | Het individuele niveau

3.1 Inleiding

De leerling en het gezin nemen voor de verklaring van verschillen in prestaties een cruciale positie in. De vraag waarom er verschillen zijn in prestaties naar sociaal-economische status of andere structurele kenmerken is slechts gedeeltelijk beantwoord. Bernstein (1971), Kohn (1977²), Bourdieu (1977) en anderen hebben licht op deze vraag geworpen, maar de empirische evidentie voor hun theorieën ontbreekt gedeeltelijk of geheel. De verklaring van verschillen in prestaties naar sociaal milieu is de laatste decennia dan ook nauwelijks verfijnd (vergelijk Dronkers, 1990; Peschar & Wesselingh, 1995).



Om verschillen in prestaties tussen leerlingen te verklaren, gebruik ik de structuur van concentrische cirkels, zie de bovenstaande figuur. In hoofdstuk 1 heb ik uitgelegd dat deze structuur uit zes cirkels bestaat, te weten: (5) structurele

kenmerken van ouders en leerlingen; (4) interactie tussen ouders en kinderen in het algemeen; (3) interactie tussen ouders en kinderen over schoolse zaken; (2) algemene kenmerken van leerlingen; (1) schoolse kenmerken van leerlingen; (0) prestaties van leerlingen.

Met behulp van deze structuur beoog ik meer inzicht te geven in de relaties tussen ouder- en leerlingkenmerken. Ook lijkt het mogelijk om hiermee de effecten van kenmerken als de sociaal-economische status of de etniciteit op prestaties te verklaren door intermediaire variabelen. Bij het opstellen van deze structuur ben ik beïnvloed door Ryan en Adams (1995). Zij stellen voor om de verschillende invloeden op leerprestaties cirkelsgewijs in te delen, teneinde de kennis over effecten op leerprestaties te verfijnen (vergelijk Bierman, 1996; Paterson, 1992).

In dit hoofdstuk ga ik de invloed op prestaties na van zowel structurele als culturele kenmerken. In paragraaf 3.2 geef ik de structuur van concentrische cirkels nadere invulling.ⁱ Om deze structuur empirisch te onderbouwen, gebruik ik gegevens afkomstig uit het cohortonderzoek VOCL'93. In paragraaf 3.3 beschrijf ik deze dataset, operationaliseer ik de leerling- en gezinskenmerken en geef ik beschrijvende statistieken. Vervolgens staan in paragraaf 3.4 de resultaten van de multiniveau analyse naar de invloed van leerling- en gezinskenmerken op de prestaties Nederlands en wiskunde centraal. De onderzoeksvraag hierbij luidt: *In hoeverre en waarom zijn kenmerken op individueel niveau van invloed op de prestaties van leerlingen in het derde leerjaar van het voortgezet onderwijs?*

3.2 Belang van leerling- en gezinskenmerken

In deze paragraaf bespreek ik structurele en culturele kenmerken die op individueel niveau van invloed kunnen zijn op de prestaties van leerlingen. Deze bespreking is niet uitputtend. Ik heb er voor gekozen om alleen kenmerken te bespreken waarover in de dataset gegevens beschikbaar zijn. Alle kenmerken die ik behandel, komen dus terug in paragraaf 3.3 en 3.4. Door deze pragmatische keuze ontbreekt een aantal relevante kenmerken in deze bespreking. Om dit verzuim tegen te gaan, heb ik een aantal keren in een noot het belang van een bepaald leerling- of gezinskenmerk toegelicht. Voor de keuze van de kenmerken heb ik geen gebruik kunnen maken van een overzicht met effectieve leerling- en gezinskenmerken. Een dergelijk overzicht ontbreekt namelijk. Resultaten van talrijke onderzoeken combinerend, probeer ik met deze invulling van de structuur van concentrische cirkels daartoe een aanzet te geven.

Structurele leerling- en gezinskenmerken (cirkel 5)

Allereerst bespreek ik het belang van verschillende structurele kenmerken. In deze bespreking behandel ik de volgende kenmerken: de sociaal-economische status, de etniciteit, de gezinssamenstelling en de sekse.

Overduidelijk is een effect aangetoond van de *sociaal-economische status* op de prestaties van leerlingen. Onderzoekers beargumenteren op verschillende manieren hoe de invloed van de sociaal-economische status verloopt (Dronkers, 1986). Allereerst wijzen ze op de invloed van de *financiële positie* van ouders. Ouders met een hoog inkomen hebben meer mogelijkheden om voorwaarden te scheppen waaronder hun kinderen goede prestaties kunnen leveren. In de tweede plaats kan worden gekeken naar de *arbeidsomstandigheden* van ouders. Naarmate het beroepsniveau hoger is, beschikt iemand doorgaans over meer autonomie. De mate van autonomie waarover een kostwinner beschikt hangt samen met andere kenmerken van die persoon, zoals het leefpatroon, de opvoedingsstijl en de ambities voor zichzelf en zijn of haar kinderen. In de derde plaats zijn aspecten van het *opleidingsniveau* van ouders van belang. In het algemeen zijn ouders die lager zijn opgeleid beperkter in hun mogelijkheden. Ze kunnen hun kind bijvoorbeeld minder goed ondersteunen met huiswerk, zijn minder bekend met de prestaties die hun kind op school moet leveren, zijn minder goed in staat om met docenten te praten over leervorderingen of gedragsproblemen, beschikken over minder cultureel kapitaal en geven in de opvoeding minder intellectuele prikkels dan ouders met een hoge opleiding. Het effect van de sociaal-economische status van ouders op de prestaties van hun kind kan verlopen via andere kenmerken, zoals het onderwijsondersteunend gedrag, het leefpatroon en de opvoedingsstijl van ouders. Ook kan de relatie tussen sociaal milieu en prestaties verlopen via de intelligentie van leerlingen. Dronkers (1997a) stelt overigens dat de opleiding en het beroep van ouders alleen nog een invloed hebben op de startpositie van leerlingen in het voortgezet onderwijs. De verschillen die bij aanvang aanwezig zijn nemen in het voortgezet onderwijs niet toe.

Verschillen in *eticiteit* kunnen ook van betekenis zijn voor de prestaties van leerlingen. De situatie waarin leerlingen uit de ene etnische groep zich bevinden, kan veel gunstiger zijn dan die van andere etnische groepen. In het algemeen presteren autochtone leerlingen beter dan allochtone leerlingen (Driessen, 1995; De Jong, 1989). De samenhang tussen sociaal-economische status en etniciteit is echter groot. Wanneer er rekening wordt gehouden met het ene kenmerk dan is er in sommige onderzoeken geen of een klein effect van het andere kenmerk op de prestaties van leerlingen. In het merendeel van de Nederlandse onderzoeken zijn echter beide kenmerken van invloed, waarbij volgens Driessen (1995) de teneur is dat de sociaal-economische status belangrijker is dan de etniciteit. Behalve van de

gemiddeld genomen lagere sociaal-economische status van allochtone leerlingen, kan van de taal die thuis wordt gesproken en de verblijfsduur in Nederland een effect uitgaan op de prestaties van leerlingen (Hacquebord, 1989). Ook is het mogelijk dat prestatieverschillen tussen leerlingen van verschillende etnische afkomst te herleiden zijn tot de cultuur of de migratiegeschiedenis van een groep (Dronkers, 1997a; Ledoux, 1996).

Uit kwalitatief onderzoek van Crul (1996) blijkt dat Marokkaanse en Turkse ouders door gebrek aan kennis en inzicht in het Nederlandse onderwijs geen hulp kunnen geven bij de leerstof en ook niet kunnen adviseren over vakkenpakketten en schoolkeuzen. Deze ouders kunnen hun kinderen alleen ondersteunen door huiswerk en schoolresultaten te controleren, door het aanmoedigen tot prestaties of het geven van morele steun.ⁱⁱ

Prestaties van leerlingen kunnen ook samenhangen met de *gezinssamenstelling*. Zo sluiten kinderen uit eenoudergezinnen hun schoolloopbaan op een lager niveau af, ook wanneer onderzoekers controleren voor andere relevante achtergrondkenmerken, zoals schoolprestaties en sociaal-economische status (Bosman, 1993; Downey, 1994). De band tussen moeder en kind is in moedergezinnen volgens Bosman vaak te hecht, waardoor er onduidelijke gezagsverhoudingen kunnen ontstaan. Bovendien kan in deze gezinnen de regelcapaciteit onder druk staan, waardoor de opvoedingsstijl meer permissief is. Ook wijst ze voor deze gezinnen op de invloed van de verslechterde inkomenspositie, het minder sociaal-aangepaste gedrag van de kinderen en de opvoedingsonzekerheid. De ontstaanswijze van eenoudergezinnen is hierbij waarschijnlijk ook belangrijk. De gevolgen van een echtscheiding blijken namelijk groter dan die van de dood van een van de ouders. Vooral de spanningen die vaak met een echtscheiding gepaard gaan, kunnen nadelig werken op kinderen (vergelijk Dronkers, 1997c). Behalve het aantal ouders kan ook het aantal kinderen in een gezin van invloed zijn op de prestaties van een leerling, zo blijkt uit een ander onderzoek van Downey (1995). Naarmate het kindertal toeneemt, daalt namelijk de gemiddelde contacttijd per kind en daarmee het profijt van ouderlijke hulpbronnen. In grotere gezinnen praten kinderen volgens Downey minder vaak met hun ouders, zijn de ouderlijke verwachtingen over de loopbaan van hun kinderen minder hoog en leggen ouders minder geld opzij voor het doorleren van een kind. Het aantal activiteiten dat ouders met hun kinderen ondernemen neemt eveneens af naarmate de gezinsgrootte toeneemt. Ook ontbreekt in grote gezinnen vaak een plek voor kinderen om te leren. Voor deze ouderlijke hulpbronnen geldt dat ze per kind geld kosten of dat ze niet of moeilijk kunnen worden gedeeld met broers of zussen. Het aantal kinderen heeft ook een effect op leerprestaties, wanneer wordt gecontroleerd voor onder meer het opleidingsniveau van ouders, het beroepsniveau van de kostwinner en het inkomen van het huishouden (Downey, 1995). Ook de

positie in de kinderrij kan van invloed zijn op de leerprestaties van een kind. Gemiddeld genomen presteren oudste en enige kinderen het beste, zo blijkt uit een overzicht van Falbo en Polit (1986). Enige kinderen komen echter steeds vaker uit eenoudergezinnen en zouden het om die reden tegenwoordig minder goed kunnen doen in het onderwijs.

Bij aanvang van het voortgezet onderwijs verschillen prestaties naar *seks* (Hustinx, 1996; Simons, 1995). Meisjes presteren gemiddeld genomen beter bij taal en jongens doen het beter bij rekenen. Meisjes kiezen voor iets hogere vormen van voortgezet onderwijs, maar daarbinnen voor minder exacte vakkenpakketten. Op het eindexamen halen ze gemiddeld genomen lagere cijfers. Bij deze slechtere examenresultaten moet echter in aanmerking worden genomen dat meisjes gemiddeld genomen het examen op een hoger niveau afleggen dan jongens. Voor een verklaring van prestatieverschillen tussen jongens en meisjes kan worden gedacht aan mogelijke verschillen tussen jongens en meisjes in prestatiemotivatie, inzet, leefpatroon en schoolbeleving. Verder laten zowel Ho en Willms (1996) als Ainsworth-Darnell en Roscigno (1997) zien dat ouders in de Verenigde Staten vaker met meisjes over school praten dan met jongens. Wanneer ouders echter met de school contact opnemen dan gaan de gesprekken vaker over hun zonen dan over hun dochters.

Algemene interactie in het gezin (cirkel 4)

In de bespreking van de algemene interactie in het gezin komen twee kenmerken aan bod, te weten de opvoedingsstijl en het aantal contacturen tussen ouders en kinderen.ⁱⁱⁱ

De interactie in het gezin wordt door de *opvoedingsstijl* beïnvloed. Vooral de zogeheten autoritatieve opvoedingsstijl heeft een positieve invloed op de leerprestaties van kinderen (Steinberg, Elmen & Mounts, 1989; Steinberg, Lamborn, Dornbusch & Darling, 1992). Kenmerken van deze opvoedingsstijl zijn toezicht, ondersteuning en autonomie. Ouders oefenen controle uit, maar moedigen ook het zelfstandig gedrag van hun kinderen aan. Volgens Steinberg e.a. (1989; 1992) verloopt de invloed van een autoritatieve opvoedingsstijl via de betrokkenheid van ouders bij schoolse zaken en via de inzet, de prestatiemotivatie en het zelfbeeld van het kind.

Ook zijn er aanwijzingen voor een effect van het *aantal contacturen* tussen ouders en kinderen. Kinderen bij wie overdag niemand thuis is maken minder vaak huiswerk en hebben in het algemeen meer moeite om alles op tijd af te krijgen. Volgens Coleman en Hoffer (1987, p. 223) kunnen kinderen wanneer ze weinig contact met hun ouders hebben, onvoldoende profiteren van hun kennis en vaardigheden. “(O)utcomes for children are strongly affected by the human capital

possessed by their parents. But this human capital can be irrelevant to outcomes for children if parents are not an important part of their children's lives, if their human capital is employed exclusively at work or elsewhere outside the home.” Uit onderzoek van Muller en Kerbow (1993) blijkt dat naarmate kinderen vaker alleen thuis zijn, ze slechter presteren.

Schoolse interactie in het gezin (cirkel 3)

Van de schoolse gezinskenmerken behandel ik hier één kenmerk, namelijk het onderwijsondersteunend gedrag van ouders. Voor de relatie tussen onderwijsondersteunend gedrag door ouders en prestaties van leerlingen geldt dat het uiterst moeilijk is om na te gaan wat oorzaak en gevolg is. In empirisch onderzoek is bij de interpretatie van dergelijke verbanden dus voorzichtigheid geboden.

Om de prestaties van hun kind te bevorderen kunnen ouders tal van activiteiten ondernemen. Ze kunnen bijvoorbeeld met hun kind praten over schoolse zaken of helpen bij het maken van huiswerk. Het blijkt dat de relatie tussen de meeste vormen van *onderwijsondersteunend gedrag* van ouders en de prestaties van hun kinderen negatief is (Kuyper & Guldemon, 1997; Muller & Kerbow, 1993). Naarmate de prestaties van leerlingen slechter zijn, helpen ouders vaker bij het maken van huiswerk, spreken ze hun kind eerder aan op de resultaten en nemen ze vaker contact op met school. In deze gevallen zou ik willen spreken van 'crisis-interventie'.

Het is ook denkbaar dat sommige ouders vanuit opvoedingsprincipes hun kinderen bepaalde vaardigheden of een zeker arbeidsethos willen bijbrengen. In dat geval is het waarschijnlijk dat ouders niet alleen bij slechte maar ook bij goede prestaties van hun kinderen over school praten, toezien op het huiswerk en het televisiekijken binnen bepaalde perken trachten te houden. Deze 'groei-interventie' kan gunstig zijn voor de vooruitgang in prestaties van leerlingen en kan aansluiten op het natuurlijke leerklimaat in een gezin (Simons, 1995). Wanneer kinderen thuis zien hoe volwassenen stap voor stap een bepaald probleem oplossen, nemen zij dat als het ware vanzelf over. Op die manier kunnen kinderen thuis leren leren en leren denken. Vooral in de voorschoolse fase legt dit leerklimaat gewicht in de schaal, maar ook in latere fases kan hiervan een invloed uitgaan.

Algemene kenmerken van leerlingen (cirkel 2)

Tot de algemene leerlingkenmerken reken ik hier het zelfbeeld en het leefpatroon van leerlingen.^{iv}

Het *zelfbeeld* is opgebouwd uit eigenschappen van een persoon die naar iemands eigen mening karakteristiek voor hem of haar zijn, zoals het karakter, de

vaardigheden en het uiterlijk (Bandura, 1988; Shavelson, Hubner & Stanton, 1976). Het zelfbeeld van mensen hangt samen met hun zelfvertrouwen. Personen met een positief zelfbeeld, zo luidt de attributietheorie van Weiner (1979), gaan ervan uit dat zij een grote kans hebben om een taak succesvol te volbrengen. Mocht die taak niet lukken, dan zullen ze hun falen toeschrijven aan bijvoorbeeld pech of een slecht humeur. Personen met een positief zelfbeeld zoeken de oorzaak van het falen in instabiele kenmerken en zullen daarom hun verwachtingen voor het volbrengen van een soortgelijke taak niet bijstellen. Succes schrijven ze daarentegen toe aan stabiele kenmerken, zoals hun geschiktheid en doorzettingsvermogen, waardoor hun verwachtingen voor toekomstig succes worden verhoogd en het positieve zelfbeeld wordt bevestigd. Omgekeerd schrijven personen met een negatief zelfbeeld volgens Weiner falen toe aan stabiele kenmerken en succes aan instabiele kenmerken.

Ook van het *leefpatroon* van leerlingen buiten schooltijd kan een invloed uitgaan op hun prestaties. Het vrijetijdsgedrag van leerlingen concurreert als het ware met de inzet voor huiswerk. Op straat rondhangen met vrienden en uitgaan geven een aanduiding voor een leefpatroon waarin niet-schoolse zaken een belangrijke plaats innemen. Talrijke scholieren besteden tijd aan dergelijke activiteiten (Steinberg, 1996). Ook een bijbaantje kan veel tijd opslokken. Wanneer een leerling te veel tijd besteedt aan uitgaan of aan een bijbaantje kan dat ook tot oververmoeidheid leiden. Steinberg rapporteert dat scholieren die meer dan twintig uur in de week werken minder huiswerk maken, meer spijbelen, een lagere schoolbeleving hebben, minder hoge onderwijsaspiraties hebben en lagere cijfers behalen. Andere vrijetijdsgedragingen, zoals lezen, zijn wellicht gunstig voor hun prestaties, met name voor hun taalprestaties. Het leefpatroon van leerlingen zou wel eens een belangrijke verklaring kunnen bieden voor verschillen in prestaties, omdat de invloed van vrienden voor scholieren in het voortgezet onderwijs belangrijk is.^v

Schoolse kenmerken van leerlingen (cirkel 1)

Tot slot bespreek ik het belang van de schoolse leerlingkenmerken. In deze bespreking ga ik in op de volgende kenmerken: eerdere prestaties, intelligentie, inzet, werkhouding, prestatiemotivatie en schoolbeleving.^{vi}

De *eerdere prestaties* van leerlingen verklaren uiteraard voor een belangrijk deel de latere verrichtingen. De eerdere prestaties staan niet gelijk aan de intelligentie van een leerling. De overerfbaarheid van leerprestaties is namelijk een stuk kleiner dan die van intelligentie (Jensen, 1969, p. 59). “The fact that scholastic achievement is considerably less heritable than intelligence also means that many other traits, habits, attitudes, and values enter into a child's performance in school besides just his intelligence, and these non-cognitive factors are largely environmentally determined, mainly through influences within the child's family.

This means there is potentially much more we can do to improve school performance through environmental means than we can do to change intelligence per se.” Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, Halpern, Loehlin, Perloff, Sternberg en Urbina (1996) wijzen in dit opzicht op de wiskundeprestaties van leerlingen in Japan en Taiwan. Zij schrijven de goede prestaties van deze Aziatische leerlingen vooral toe aan verschillen in cultuur en scholing en minder aan verschillen in intelligentie.^{vii}

Onderzoekers stellen een sterke invloed van *intelligentie* op leerprestaties vast (Fraser, Walberg, Welch & Hattie, 1987). Neisser e.a. (1996) geven een overzicht van wat bekend en onbekend is over intelligentie. Zij stellen dat de variantie in intelligentie bij kinderen voor 45 procent en bij jongeren en volwassenen voor 75 procent door erfelijke factoren wordt bepaald. Naarmate mensen opgroeien zijn verschillen in intelligentie meer en meer toe te schrijven aan genotypische, erfelijke verschillen. Zij concluderen dat erfelijke en omgevingskenmerken van invloed zijn op de intelligentie, maar dat het tamelijk onbekend is hoe die invloed zich uit.

De mate van *inzet* van leerlingen voor school is van belang voor hun prestaties. Keith (1982) laat zien dat naarmate leerlingen meer tijd aan huiswerk besteden, gegeven hun intelligentie, hun cijfers hoger worden. In de praktijk zijn het lang niet altijd de leerlingen met de meeste aanleg die hard werken en veel huiswerk maken. Het zijn juist de wat minder intelligente leerlingen die de negatieve effecten van hun lagere intelligentie trachten te compenseren met een verlenging van de huiswerktijd. Hierbij passen veel leerlingen een optimaliseringsstrategie toe (Kuyper & Swint, 1996). Met niet al te veel inspanning willen ze een 'veilige' voldoende behalen.

De *werkhouding* van leerlingen is eveneens relevant. Zo hebben veel leerlingen niet de rust en discipline om op te letten. Ook vertonen leerlingen geregeld uitstelgedrag, wat inhoudt dat het ze moeite kost om aan het huiswerk te beginnen en dat ze deze activiteit het liefst zo ver mogelijk voor zich uitschuiven. Dit gedrag correleert negatief met de gemiddelde rapportcijfers van leerlingen (Kuyper & Swint, 1996).

De gerichtheid van leerlingen op het nastreven van succes en de daaraan verbonden positieve gevoelens zijn ook van invloed op hun prestaties (Fraser e.a., 1987; Steinberg, Elmen & Mounts, 1989). Kuyper en Guldemon (1997) laten zien dat de *prestatiemotivatie* van havo- en vwo-leerlingen in leerjaar drie relatief hoog correleert met hun prestaties in leerjaar vijf. Ook hangt de prestatiemotivatie sterk samen met de hoeveelheid tijd die leerlingen aan hun huiswerk besteden.

De *schoolbeleving* van leerlingen hangt samen met verscheidene aspecten, zoals de opvatting over schoolregels, de relatie met docenten en klasgenoten. Uit onderzoek van Van der Linden en Dijkman (1989) onder vijfhonderd jongeren van

12 tot 21 jaar blijkt dat jongeren vooral positief oordelen over boeiende vakken, goed contact met leeftijdsgenoten, een gezellige sfeer op school, goed contact met bepaalde docenten en aantrekkelijke buitenschoolse activiteiten. Negatief oordelen jongeren over strenge regels en te veel huiswerk. De waardering van leerlingen over een docent als persoon is gemiddeld vrij gunstig. Over de docent als didacticus oordelen leerlingen negatiever dan over de docent als persoon. Het oordeel over de didactische kwaliteiten van docenten is gemiddeld gunstig noch ongunstig (Stoel, 1980). Hermans (1981) komt tot de conclusie dat leerlingen de afstand van een docent tot hun leefwereld als groot ervaren. De bedoelingen van docenten zijn volgens leerlingen goed. Zo meent 80 procent van de leerlingen dat docenten bereid zijn om hen te helpen en dat er talrijke mogelijkheden tot contact zijn, maar leerlingen zoeken bij problemen zelden ondersteuning van docenten. Voor de waardering van leerlingen over hun docenten is het belangrijk dat docenten hen serieus nemen en verantwoordelijkheden geven (Coleman, Hoffer & Kilgore, 1982; Hermans, 1981; Rutter, Maughan, Mortimore en Ouston, 1979). Ook is het van gewicht dat docenten betrokken zijn bij leerlingen. Volgens Bryk, Lee en Holland (1993, pp. 99-100) werkt de sterke betrokkenheid van docenten bij leerlingen positief op hen door. "This broad concern and extended teacher role was not lost on students, who saw their instructors as interested in them, as patient and understanding, yet also as firm and committed to high standards." Over de relatie met hun klasgenoten zijn leerlingen positiever dan over de omgang met docenten (Hermans, 1981; Kuiper & Schuurman, 1981; Stoel, 1980). Het oordeel over de omgang met medeleerlingen is echter niet voor iedereen positief. Verscheidene leerlingen vinden de sfeer op school niet bijzonder, omdat leerlingen elkaar bijvoorbeeld niet in hun waarde laten. Zo zijn leerlingen soms bang voor onverwachte vragen en beurten, omdat ze denken dat ze dan af zullen gaan voor hun klasgenoten. Rutter e.a. (1979) wijzen er op dat de mate waarin leerlingen zich thuis voelen op school van invloed is op het gedrag en de examenresultaten van leerlingen.

Besluit

De empirische steun voor de invloed van structurele kenmerken op de prestaties van leerlingen is stevig. De effecten van de sociaal-economische status, de etniciteit en de sekse zijn onomstreden. Ook voor de invloed van de gezinssamenstelling is empirisch steun aanwezig. Wat de schoolse en algemene leerlingkenmerken betreft vinden onderzoekers herhaaldelijk effecten van de intelligentie, de inzet, de prestatiemotivatie, het zelfbeeld, de schoolbeleving en het leefpatroon. Ook voor culturele gezinskenmerken als het onderwijsondersteunend gedrag en de opvoedingsstijl van ouders is steun aanwezig. Voor zover bekend verschillen de

uitkomsten van Nederlands onderzoek naar leerling- en gezinskenmerken niet van die van buitenlandse studies, met name wanneer het de richting van effecten betreft.

Uit bovenstaand overzicht kan een aantal verbanden over de *structurele leerling- en gezinskenmerken* worden afgeleid. De verwachting is dat de score op de toetsen Nederlands en wiskunde hoger is voor: (1) autochtone leerlingen; (2) kinderen uit een tweeoudergezin; (3) oudste en enige kinderen. De verwachting is dat (4a) meisjes op Nederlands beter presteren dan jongens en dat (4b) jongens op wiskunde beter presteren dan meisjes. Verder is de voorspelling dat de score op beide toetsen toeneemt naarmate een leerling uit een gezin komt: (5) met een hogere sociaal-economische status; (6) met minder kinderen. De *algemene en schoolse interactie in het gezin* zal ook de prestaties beïnvloeden. De voorspelling is dat de score op beide toetsen hoger is naarmate: (7) de opvoedingsstijl meer autoritatief is; (8) leerlingen meer contacturen hebben met hun ouders; (9) leerlingen minder vaak door hun ouders worden aangespoord tot of gecontroleerd op het maken van huiswerk (crisis-interventie); (10) leerlingen vaker met hun ouders praten (groei-interventie). Van de *algemene kenmerken van leerlingen* kunnen bepaalde kenmerken de leerprestaties bevorderen, maar andere kenmerken kunnen die prestaties in de weg zitten. De verwachting is dat de score op de toetsen hoger is naarmate een leerling: (11) minder vrije tijd besteedt op straat met vrienden; (12) minder vrije tijd besteedt aan een bijbaantje; (13) minder vrije tijd besteedt aan uitgaan; (14) een positiever zelfbeeld heeft. Voor de score op Nederlands wordt verwacht dat (15) leerlingen hoger score naarmate ze meer vrije tijd besteden aan lezen en hobby's. Tot slot zijn de *schoolse kenmerken van leerlingen* van belang voor de leerprestaties. De verwachting is dat de score op de toetsen hoger is naarmate leerlingen: (16) positiever oordelen over hun medeleerlingen; (17) positiever oordelen over hun docenten en de schoolregels; (18) een hogere prestatiemotivatie hebben; (19) het maken van huiswerk minder uitstellen; (20) meer dagen in de week huiswerk maken; (21) hoger scoorden op de non-verbale intelligentietests; (22) hoger scoorden op de informatieverwerking-, taal- en rekentoets uit leerjaar één.

3.3 Meten van leerling- en gezinskenmerken

In Nederland is er veelvuldig onderzoek gedaan naar de school- en beroepsloopbanen van leerlingen. Om de effecten van leerling-, gezins-, docent- of schoolkenmerken op prestaties te kunnen onderzoeken, zijn er veel gegevens nodig. Sinds het einde van de jaren vijftig (Matthijssen & Sonnemans, 1959) zijn er daarom verscheidene cohorten met gegevens over loopbanen van leerlingen aangelegd. In dit onderzoek wordt het meest recente cohort van leerlingen uit het

voortgezet onderwijs gebruikt, namelijk VOCL'93 (Voortgezet Onderwijs Cohort Leerlingen). Deze gegevens worden in eerste instantie verzameld voor de evaluatie van de basisvorming, maar zijn ook geschikt om verbanden te leggen tussen de achtergrond en de leeromgeving van leerlingen met hun prestaties.

In deze paragraaf ga ik allereerst in op de dataverzameling en representativiteit van VOCL'93. Vervolgens operationaliseer ik de kenmerken van leerlingen en gezinnen en tot slot geef ik de beschrijvende statistieken voor deze variabelen.

Een groot databestand: VOCL'93

Voor VOCL'93 heeft het CBS een steekproef van 770 scholen getrokken (Brandsma, Lugthart & Van der Werf, 1997). Van deze 770 scholen hebben 333 hun medewerking verleend. De respons bedroeg dus 43 procent. Vervolgens is niet alleen een steekproef van scholen maar ook van klassen getrokken. Van scholen met meer dan één klas in een leerjaar is ongeveer de helft van die klassen getrokken. Uiteindelijk resulteerde dit in een steekproef van 20.331 leerlingen binnen 880 klassen binnen 333 scholen.

Op deze scholen hebben het CBS, het GION en het OCTO in het voorjaar van 1994 gegevens verzameld bij leerlingen uit het eerste leerjaar van het voortgezet onderwijs, hun docenten en de directie. Het CBS heeft in 1994 ook gegevens verzameld bij ouders. In deze vragenlijst werd onder meer naar de sociaal-economische status, de etniciteit en de gezinssamenstelling gevraagd. De respons op de oudervragenlijst bedroeg 85 procent.^{viii}

Tabel 3.1: Respons VOCL'93

	SCHOLEN	KLASSEN	LEERLINGEN
Steekproef CBS	770	onbekend	onbekend
Deelname in 1993-'94	333	880	20.331
Deelname in 1995-'96	261	614	9.766
- Toetsscore Nederlands of wiskunde	258	604	9.469
- LKS-vragenlijst	150	453	7.214
Bestand voor analyses	149	452	6.896

Vlak voor en na de zomervakantie van 1996 zijn er wederom gegevens verzameld bij leerlingen die in het schooljaar 1995-'96 in het derde leerjaar zaten. Het ging hierbij om 9.766 leerlingen binnen 614 klassen binnen 261 scholen. Ook zijn er

opnieuw gegevens verzameld bij de ouders, docenten en directieleden van de leerlingen.^{ix}

In 1995-'96 deden nog 261 van de 333 scholen mee. Het aantal scholen dat serieus heeft meegewerkt aan de meting op het tweede tijdstip bedraagt echter minder dan 200.^x Van de 261 scholen zijn er 44 waar slechts één of twee leerlingen een toets hebben gemaakt. Voor 30 procent van de scholen geldt dat er in het schooljaar 1995-'96 minder dan vijftien leerlingen aan de toetsen hebben deelgenomen. Twee jaar eerder gold dit voor 1 procent van de scholen.

Tegelijkertijd met de reguliere vragenlijsten uit VOCL'93 is er voor dit onderzoek een aanvullende vragenlijst afgenomen onder de leerlingen uit het cohortonderzoek, de LKS (Leerlingen-Klassen-Scholen)-vragenlijst. Van de 9.766 leerlingen die op het tweede meetmoment aan VOCL'93 hebben deelgenomen, hebben 7.214 leerlingen aanvullende informatie gegeven.^{xi} De respons op deze aanvullende vragenlijst bedroeg 74 procent. In deze aanvullende vragenlijst is gevraagd naar culturele leerling- en gezinskenmerken, zoals de invloed die ouders en kinderen hebben op beslissingen, de tijdsbesteding van leerlingen na schooltijd, het zelfbeeld van leerlingen, de mening van leerlingen over hun docenten, medeleerlingen en de regels op school. Ook voor andere onderzoeken zijn er vragen gesteld aan de leerlingen, bijvoorbeeld over hun gelovige achtergrond (zie: Veenstra & Dijkstra, 1998).

Het bestand dat ik gebruik voor analyses bevat 6.896 leerlingen binnen 452 klassen binnen 149 scholen.^{xii} Er zijn dus nog 318 leerlingen verwijderd uit het bestand. Hierbij gaat het om leerlingen van een school waarop alle leerlingen een te sterk afwijkende score hebben op de toetsen en om leerlingen met extreem lage scores op de toets Nederlands.^{xiii} Deze steekproef van scholen, klassen en leerlingen is qua grootte geschikt voor multiniveau analyse. De gegevens over de respons zijn samengevat in Tabel 3.1. In totaal zijn in dit onderzoek 12.435 toetsscores Nederlands en wiskunde gebruikt. De geneste structuur kan daarom nog worden uitgebreid door te stellen dat er 12.435 toetsscores binnen 6.896 leerlingen zijn.

Representativiteit

De 333 scholen die in het schooljaar 1993-'94 meededen aan VOCL'93 vormen naar schooltype, denominatie, regio en urbanisatiegraad een representatieve steekproef van alle scholen voor voortgezet onderwijs in Nederland (Brandsma, Lugthart & Van der Werf, 1997), zie Bijlage 1. Voor de LKS-steekproef geldt dat de steekproef niet representatief is naar regio en urbanisatiegraad. De uitval op de tweede meting deed zich sterker voor bij enerzijds scholen uit het westen en het oosten van Nederland en anderzijds scholen uit plaatsen met meer dan

honderdduizend inwoners, zie opnieuw Bijlage 1. Aangezien dit onderzoek vooral gericht is op verklaren en niet zo zeer op beschrijven, is deze vertekening niet problematisch.

Voor de leerlingen die in de analyses worden meegenomen geldt dat ze onvertraagd zijn doorgestroomd naar het derde leerjaar. Bij leerlingen die in de eerste of tweede klas zijn blijven zitten zijn op het tweede tijdstip namelijk geen gegevens verzameld. Het gaat hier dus om een enigszins selecte groep. Deze selectie is overigens gering. Voor VOCL'93 is een apart onderzoek onder vertraagden zelfs niet doorgegaan, omdat bleek dat scholen voor voortgezet onderwijs tegenwoordig in de eerste twee leerjaren nauwelijks nog leerlingen laten zitten. Begin jaren negentig bleef nog 5 procent van de leerlingen zitten in leerjaar één en 8 procent in leerjaar twee. Halverwege de jaren negentig waren die percentages ongeveer gehalveerd tot respectievelijk 3 en 4 procent (CBS, 1998).^{xiv}

Operationalisatie

Allereerst zijn er de structurele leerling- en gezinskenmerken. De gegevens hierover zijn afkomstig van de schooladministratie en de oudervragenlijst (afgenomen in schooljaar 1993-'94). In de analyses neem ik de *seks*, de *positie in de kinderrij* (enig kind; oudste maar niet enig kind; overige kinderen) en de *gezinssamenstelling* (al dan niet een tweoudergezin en het kindertal) mee. In navolging van Downey (1995) heb ik het aantal kinderen (X) omgezet in een inverse score ($1/X$), omdat de verwachting is dat het effect van het aantal kinderen op de contacttijd van ouders met elk kind non-lineair afneemt. Voor de inverse score geldt dat één versus twee kinderen een groter verschil geeft dan bijvoorbeeld vier versus vijf kinderen.

Bij de *opleiding van ouders* onderscheid ik vijf eindniveaus, te weten: (1) basisonderwijs; (2) lagere vormen van voortgezet onderwijs (lager beroepsonderwijs, mavo of de onderbouw van havo en vwo); (3) hogere vormen van voortgezet onderwijs en middelbaar beroepsonderwijs; (4) universitair en hoger beroepsonderwijs; (5) postdoctorale universitaire opleidingen (inclusief doctoraal/ingenieurs-examen oude stijl). De categorie 'minder dan basisonderwijs' komt in VOCL'93 niet meer voor, in tegenstelling tot in VOCL'89. Bij het *beroepsniveau van de kostwinner* maak ik gebruik van dezelfde indeling als in het vorige hoofdstuk: (1) mensen zonder betaald werk; (2) arbeiders; (3) zelfstandigen, landbouwers en lagere employés; (4) middelbare employés; (5) hogere employés en vrije beroepen. Door de scholen is informatie verstrekt over het geboorteland en de nationaliteit van leerlingen. Bovendien is aan de ouders gevraagd wat hun geboorteland en nationaliteit is. Met deze gegevens bepaal ik de *etniciteit* van leerlingen. Net als in het vorige hoofdstuk definieer ik een leerling als allochtoon als ten minste één van de ouders in een ander land dan Nederland is geboren.

De culturele leerlingkenmerken zijn afkomstig uit de vragenlijst voor leerlingen (met uitzondering van de eerdere prestaties en de non-verbale intelligentietests afgenomen in schooljaar 1995-'96). Om na te gaan in hoeverre leerlingen op schooldagen met volwassenen in huis zijn, zijn twee vragen gebruikt waarmee leerlingen aangeven hoeveel uren ze na schooltijd thuis zijn en welk deel daarvan zonder volwassenen. Uit deze vragen kan de tijd worden berekend die leerlingen thuis met volwassenen doorbrengen. Het *aantal contacturen* kan lopen van nul tot negen uur. Om de *opvoedingsstijl* van ouders te bepalen is een index van acht vragen gebruikt, zie Bijlage 6. Deze index is geschaald van 0 tot 100.

Om het *onderwijsondersteunend gedrag* te bepalen, gebruik ik vragen over hoe vaak leerlingen door hun ouders worden aangespoord tot en gecontroleerd op het maken van huiswerk. Verder gebruik ik een schaal bestaande uit zeven items waarin wordt gevraagd naar het aantal keren dat ouders en kinderen met elkaar praten over schoolse zaken ($\alpha = 0,77$), zie Bijlage 6.

Het *zelfbeeld* is een schaal bestaande uit elf vierpunts-items ($\alpha = 0,75$), zie Bijlage 6. Wat het *leefpatroon* van leerlingen betreft ga ik in op de tijdsbesteding aan de activiteiten 'uitgaan', 'een bijbaantje', 'op straat met vrienden' en 'lezen en hobby's', omdat ik alleen van deze kenmerken een effect op de score Nederlands verwacht. In Bijlage 6 wordt aangegeven hoe deze gegevens zijn verzameld.

Voor de operationalisatie van de schoolse kenmerken van leerlingen gebruik ik gegevens van zowel toetsen als vragenlijsten. Om de *eerdere prestaties* van leerlingen te bepalen, gebruik ik hun scores op de toetsen taal en informatieverwerking die afgenomen zijn in het eerste leerjaar. De maximale score op de taaltoets is negentien punten, met een betrouwbaarheid van 0,76. Op de informatieverwerking- en de rekentoets kunnen twintig punten worden gescoord. De Cronbachs alfa van deze toetsen is 0,78 en 0,84. Brandsma en Van der Werf (1997) beschrijven deze toetsen. Verder houd ik rekening met de *non-verbale intelligentie* van leerlingen, die in leerjaar één is gemeten met de PSB-3 test (redeneren) en de PSB-8 test (abstraheren). Horn (1969) beschrijft de psychometrische eigenschappen van deze tests, die beide uit veertig items bestaan en in acht minuten kunnen worden afgenomen. In de analyses zijn de gestandaardiseerde scores op beide tests samengenomen. De betrouwbaarheid van deze toetsen is niet bekend.

Als graadmeter voor de *inzet* van leerlingen, gebruik ik een vraag naar het aantal dagen waarop leerlingen per week huiswerk maken. Voor de *werkhouding* van leerlingen neem ik het antwoord op een vraag naar de mate waarin leerlingen het maken van huiswerk uitstellen. De *prestatiemotivatie* wordt gemeten met 21 vragen die te zamen een betrouwbare schaal vormen ($\alpha = 0,85$). Deze schaal is een bewerking van de schaal prestatiemotief uit de PMT-K-83 van Hermans (1983).

Brandsma en Van der Werf (1997) bespreken de bewerkte schaal. *Schoolbeleving* is gemeten met twee scores, namelijk het beeld dat leerlingen hebben van hun medeleerlingen en van hun docenten en de schoolregels. De mening over docenten en regels is gemeten met veertien vragen ($\alpha = 0,83$). Het beeld van medeleerlingen is de score op vier items, die geen erg betrouwbare schaal vormen ($\alpha = 0,51$). In Bijlage 6 worden beide schalen beschreven.

In dit onderzoek zijn de scores van derdeklassers op een toets *tekstbegrip Nederlands* en een toets *wiskunde* de afhankelijke variabelen. De toets Nederlands is ontwikkeld door het Cito en bestaat uit zes teksten en veertig meerkeuzevragen. De Cronbachs alfa van de schaal is 0,78. Een goed antwoord levert één punt op. De maximale score is dus veertig punten. De minimale score is op tien punten gesteld, hetgeen gelijk is aan de score die zou worden verwacht voor een leerling die alle vragen gokt. Voor wiskunde zijn er in het derde leerjaar twee toetsen gebruikt, te weten een A-versie voor mavo-, havo- en vwo-leerlingen en een B-versie voor (i)vbo-leerlingen. Beide toetsen zijn ontwikkeld door het Cito. De A-versie bestaat uit 27 open vragen en 6 meerkeuzevragen. In de B-versie zijn er 31 open vragen en 2 meerkeuzevragen. De toetsen zijn nagekeken door daartoe getrainde beoordelaars. Het aantal punten per opgave is afhankelijk van de moeilijkheidsgraad van het item en varieert van één tot acht. De maximale score voor de A- en de B-versie is respectievelijk 101 en 77 punten. Het minimale aantal punten is op 1 gesteld. Door het GION zijn met het Rasch-model de twee versies aan elkaar gelijkgeschakeld.^{xv} Na gelijkenschakeling is de betrouwbaarheid van de toetsen 0,83 voor de A- en 0,87 voor de B-versie. In dit onderzoek worden de scores na gelijkenschakeling gebruikt.

Beschrijvende statistieken

De toets Nederlands blijkt voor veel leerlingen aan de moeilijke kant. De leerlingen blijken gemiddeld genomen nauwelijks meer dan de helft van de veertig opgaven goed te beantwoorden (21,8 van de 40 vragen). In Tabel 3.2 valt te zien dat het gemiddelde toeneemt, naarmate het schooltype hoger is. De scores in deze tabel zijn gestandaardiseerd. Leerlingen in het individueel beroepsonderwijs scoren iets meer dan één standaarddeviatie onder het algemeen gemiddelde. Vwo-leerlingen scoren iets meer dan één standaarddeviatie boven dat gemiddelde.

De verschillen tussen leerlingen van een zelfde schooltype kunnen aanzienlijk zijn. Een mavo-leerling met een score van één standaarddeviatie (dat wil zeggen een binnen-schooltype standaarddeviatie) boven het mavo-gemiddelde presteert op hetzelfde niveau als de gemiddelde havo/vwo-leerling. Een havo-leerling met een score van één standaarddeviatie onder het havo-gemiddelde behaalt een zelfde score als de gemiddelde vbo/mavo-leerling.

Tabel 3.2: Toetsscores naar schooltype

NEDERLANDS	GEM.	SD	MIN.	MAX.	N
IVBO	-1,16	0,62	-2,02	1,07	416
VBO	-0,64	0,76	-2,02	1,93	1.899
VBO/MAVO	-0,25	0,94	-2,02	1,76	188
MAVO (incl. MAVO/HAVO)	0,03	0,79	-1,68	2,28	2.092
HAVO	0,57	0,72	-1,16	2,62	806
HAVO/VWO	0,88	0,68	-0,82	2,45	254
VWO	1,06	0,69	-0,65	2,62	938
Totaal	0,00	1,00	-2,02	2,62	6.593
WISKUNDE	GEM.	SD	MIN.	MAX.	N
IVBO	-1,63	0,89	-4,61	1,11	391
VBO	-0,62	0,77	-4,46	3,04	1.712
VBO/MAVO	-0,20	0,71	-2,14	1,73	167
MAVO (incl. MAVO/HAVO)	0,09	0,63	-2,53	2,34	1.722
HAVO	0,57	0,54	-1,68	2,19	753
HAVO/VWO	0,91	0,62	-1,60	2,96	257
VWO	1,08	0,59	-1,06	3,74	840
Totaal	0,00	1,00	-4,61	3,74	5.842

De beide wiskundetoetsen zijn ook voor veel leerlingen tamelijk moeilijk. Mavo-, havo- en vwo leerlingen behalen gemiddeld genomen 58 van de 101 punten en (i)vbo-leerlingen gemiddeld 41 van de 77 punten. Op de wiskundetoets lopen de scores eveneens op naarmate het schooltype hoger wordt, zie Tabel 3.2. Wederom geldt dat de verschillen tussen leerlingen van een zelfde schooltype beduidend kunnen zijn. Een havo-leerling met een score van één standaarddeviatie boven het havo-gemiddelde heeft een score die iets hoger is dan de score van de gemiddelde vwo-leerling. Omgekeerd heeft een vwo-leerling met een score van één standaarddeviatie onder het vwo-gemiddelde een score die iets lager is dan de score van de gemiddelde havo-leerling. Uit een vergelijking van de z-scores op de beide toetsen valt op dat ivbo-leerlingen slechter scoren op wiskunde dan op Nederlands. De correlatie tussen de twee toetsen uit leerjaar drie is 0,63.

Tabel 3.3: Overzicht van leerling- en gezinskenmerken

CIRKEL 1: SCHOOLSE LEERLINGKENMERKEN

Prestaties leerjaar 1	GEM.	SD	MIN.	MAX.	N
- informatieverwerking (20 items)	12,2	4,0	0	20	6.896
- taal (19 items)	11,7	3,8	0	19	6.896
- rekenen (20 items)	11,4	4,7	0	20	6.896
Non-verbale intelligentie leerjaar 1					
- redeneren (40 items)	25,4	5,2	2	40	6.894
- abstraheren (40 items)	23,8	7,0	1	40	6.894
Inzet					
- aantal dagen huiswerk	4,7	1,5	0	7	6.586
Werkhouding					
- uitstellen van huiswerk in uren	1,5	1,7	0	6	6.201
Prestatiemotivatie (21 vierpunts-items)	53,2	8,7	23	83	6.277
Schoolbeleving / mening over:					
- docenten en regels (14 vierpunts-items)	38,9	6,8	14	56	6.861
- klasgenoten (4 vierpunts-items)	11,8	2,2	4	16	6.801

CIRKEL 2: ALGEMENE LEERLINGKENMERKEN

Zelfbeeld (11 vierpunts-items)	34,4	4,9	14	44	6.838
Leefpatroon (percentages)					
- uitgaan	9,3	7,3	0	35	6.611
- bijbaan	7,3	8,2	0	35	6.601
- lezen en hobby's	8,5	6,4	0	35	6.689
- op straat met vrienden	10,2	7,0	0	35	6.579

CIRKEL 3: SCHOOLSE INTERACTIE IN HET GEZIN

Onderwijsondersteunend gedrag					
- huiswerkcontrole (2 vierpunts-items)	3,0	1,2	2	8	6.292
- praten met ouders (7 vierpunts-items)	16,0	2,8	7	21	6.798

CIRKEL 4: ALGEMENE INTERACTIE IN HET GEZIN

Aantal contacturen	4,3	2,1	0	9	6.283
Autoritaire opvoedingsstijl (percentage)	19,2	17,0	0	100	6.878

CIRKEL 5: STRUCTURELE KENMERKEN

Sekse (1=jongens)	0,49	0,50	0	1	6.896
Positie in de kinderrij					
- enig kind	0,11	0,32	0	1	6.048
- oudste kind	0,33	0,47	0	1	6.048
- overige kinderen	0,56	0,50	0	1	6.048

Gezinssamenstelling					
- aantal ouders (1=twee ouders)	0,86	0,35	0	1	6.823
- aantal kinderen	2,8	1,4	1	10	6.041
Sociaal-economische status					
- opleidingsniveau moeder	2,4	0,9	1	5	6.041
- opleidingsniveau vader	2,6	1,1	1	5	6.041
- beroepsniveau kostwinner	3,0	1,3	1	5	6.041
Etniciteit (1=allochtoon)	0,06	0,23	0	1	6.896
Geén oudervragenlijst	0,12	0,33	0	1	6.896

Voor de leerling- en gezinskenmerken heb ik de gemiddelde score, standaarddeviatie, het minimum, het maximum en het aantal cases berekend. Deze gegevens zijn samengevat in Tabel 3.3. Op de prestatietoetsen in het eerste leerjaar hebben leerlingen gemiddeld iets meer dan de helft van de items goed beantwoord. Enkele leerlingen hebben de maximale score behaald. Op de non-verbale intelligentietests hebben leerlingen gemiddeld genomen ruim de helft van de vragen goed.

Gemiddeld genomen maken leerlingen op vijf dagen in de week huiswerk. Nadat ze thuis zijn gekomen, beginnen ze gemiddeld na anderhalf uur aan dit huiswerk. De prestatiemotivatie van leerlingen is matig, het gemiddelde is namelijk vrijwel gelijk aan het schaal midden van 52,5 (de minimale score is 21 en de maximale score is 84). De mening over medeleerlingen is vrij positief. Ook de mening over docenten en regels is positief.

Het zelfbeeld van leerlingen is eveneens vrij positief. Van hun vrije tijd besteden leerlingen in leerjaar drie gemiddeld genomen 10 procent aan op straat rondhangen. Een deel van de leerlingen doet dat echter nooit, maar er zijn ook leerlingen die de helft van hun tijd zo doorbrengen. Aan lezen en hobby's, uitgaan en een bijbaantje besteden leerlingen respectievelijk 9, 9 en 7 procent van hun vrije tijd.^{xvi}

Met hun ouders praten leerlingen af en toe over schoolse zaken. Huiswerkcontrole komt in leerjaar drie niet vaak voor. De autoritatieve opvoedingsstijl komt in 19 procent van de situaties voor. Per dag zijn leerlingen gemiddeld ruim vier uur met volwassenen in huis.

Uit het overzicht van de structurele kenmerken kan worden opgemaakt dat 11 procent enig kind en 33 procent het oudste kind is, terwijl 56 procent een andere positie in de kinderrij inneemt. Gemiddeld genomen zijn er ongeveer drie kinderen in een gezin en zijn in 86 procent van de gezinnen beide ouders aanwezig. Het gemiddelde opleidingsniveau van vaders is hoger dan van moeders. Van de leerlingen is 6 procent allochtoon.

Van de vaders heeft 6,1 procent en van de moeders heeft 1,2 procent een doctoraal/ingenieursexamen oude stijl of een postdoctorale universitaire opleiding gevolgd; 13,5 procent van de vaders en 9,7 procent van de moeders heeft het onderwijs verlaten met een diploma van de universiteit of het hoger beroepsonderwijs; 39,2 procent van de vaders en 35,1 procent van de moeders heeft maximaal een havo-, vwo- of mbo-diploma; 21,4 procent van de vaders en 36,7 procent van de moeders heeft ten hoogste een lagere vorm van voortgezet onderwijs genoten; 19,8 procent van de vaders en 17,3 procent van de moeders heeft niet meer dan basisonderwijs afgemaakt. Gemiddeld genomen is het opleidingsniveau van de vader 2,7 en van de moeder 2,4 op een schaal van 1 tot 5. Voor het beroepsniveau van de kostwinner geldt dat 13,6 procent zonder betaald werk is; 27,4 procent arbeider is; 23,4 procent zelfstandige, landbouwer of lagere employé is; 21,3 procent middelbare employé is; 14,3 procent hogere employé is of een vrij beroep heeft. Gemiddeld genomen is het beroepsniveau van de kostwinner 3,0.

Ontbrekende waarden

Van de 6.896 leerlingen hebben 5.539 de toetsen Nederlands en wiskunde gemaakt. De overige leerlingen hebben een score op slechts één toets, namelijk 1.054 leerlingen op Nederlands en 303 op wiskunde. In de multivariate analyse kunnen niet alleen de leerlingen met een score op beide toetsen, maar ook de leerlingen met een score op één van de toetsen worden meegenomen.

Behalve op een afhankelijke variabele zijn er ook verscheidene ontbrekende waarden op onafhankelijke variabelen, zo blijkt uit Tabel 3.3. De gegevens van leerlingen en ouders zijn dus niet allemaal volledig. De ontbrekende waarden heb ik vervangen door *gemiddelde-score-imputatie*.^{xvii} Voor leerlingen waarvan de ouders géén vragenlijst hebben ingevuld (en waardoor gegevens ontbreken over de sociaal-economische status, de positie in de kinderrij en het aantal kinderen) heb ik deze methode bijvoorbeeld toegepast. Aangezien deze groep 12 procent van de leerlingen bedraagt en systematisch afwijkt van de overige leerlingen, heb ik een dummy toegevoegd voor een al dan niet door ouders ingevulde vragenlijst. De uitval is namelijk vooral toe te schrijven aan ouders van leerlingen uit een laag sociaal milieu en van allochtone leerlingen. Wanneer ontbrekende waarden voor de indicatoren voor de sociaal-economische status worden vervangen door de gemiddelde score van de overige leerlingen, is de vervangende waarde waarschijnlijk gemiddeld genomen te hoog. Door een dummy op te nemen kan hiervoor worden gecorrigeerd.

3.4 Effectieve leerlingen en gezinnen

Als eerste stap verricht ik een multiniveau analyse van de structurele kenmerken (cirkel vijf) op de scores Nederlands en wiskunde. Vervolgens voeg ik steeds een concentrische cirkel toe en krijg ik uiteindelijk een model met de kenmerken van alle cirkels op de toetsscores. De kenmerken uit een cirkel worden simultaan toegevoegd. Met deze volgorde hoop ik meer inzicht te geven in de relaties tussen verschillende kenmerken, waarbij ik tracht de invloed van structurele kenmerken voor een deel te verklaren door culturele kenmerken. In het model met alle cirkels schat ik de *netto* effecten van leerling- en gezinskenmerken op de scores Nederlands en wiskunde, dat wil zeggen dat er dan rekening wordt gehouden met eerdere prestaties in leerjaar één. De resultaten van de multiniveau analyse zijn weergegeven in enerzijds Tabel 3.4 en 3.5 en anderzijds Bijlage 3A en 3B.

Het doel van deze analyses is het ontrafelen van de verschillende invloeden op de prestaties van leerlingen met de structuur van concentrische cirkels. Deze analyses wijken af van analyses waarin wordt begonnen met de beste voorspeller van schoolse prestaties, bijvoorbeeld het schooladvies of de eerdere prestaties van leerlingen, en waarbij vervolgens stapsgewijs kenmerken worden toegevoegd die nog extra variantie binden. Het doel van deze analyses is namelijk niet om met zo min mogelijk variabelen zo goed mogelijk de prestaties van leerlingen te voorspellen, maar om met de structuur van concentrische cirkels de invloed van verschillende leerling- en gezinskenmerken te verklaren door intermediaire kenmerken.

Parameterschattingen

Allereerst heb ik gekeken naar de effecten van de structurele kenmerken (*cirkel vijf*) op de toetsscore Nederlands, zie het eerste model voor Nederlands in Bijlage 3A en voor wiskunde in Bijlage 3B. Alle ordinale, interval- en ratio-variabelen zijn gestandaardiseerd. Het gemiddelde is dus nul en de standaarddeviatie één. De waarden van deze variabelen zijn dan ook onderling vergelijkbaar. Voor deze variabelen zijn de regressiecoëfficiënten dus gelijk aan bèta-gewichten. De nominale variabelen, zoals de sekse of de positie in de kinderrij, zijn niet gestandaardiseerd.^{xviii}

Onder constant houding van de andere kenmerken, scoren allochtone leerlingen een vijfde standaarddeviatie lager op beide toetsen dan autochtone leerlingen. Op een zelfde wijze scoren jongens bijna een vijfde standaarddeviatie lager dan meisjes op Nederlands en bijna een kwart standaarddeviatie hoger op wiskunde. Ter vergelijking, uit Tabel 3.2 blijkt dat het gemiddelde verschil in prestaties tussen

schooltypen ongeveer een halve standaarddeviatie is. De effecten van sekse en etniciteit zijn dan ook sterk te noemen.

Oudste kinderen scoren op beide toetsen hoger en enige kinderen scoren op beide toetsen lager dan kinderen met een andere positie in de kinderrij. Leerlingen doen het beter naarmate de sociaal-economische status hoger is. Van de drie SES-indicatoren heeft het opleidingsniveau van de moeder de sterkste invloed op beide toetsscores en het beroepsniveau van de kostwinner geen invloed. Verder scoren leerlingen afkomstig uit een tweeoudergezin eentiende standaarddeviatie hoger op wiskunde dan kinderen uit één- of meeroudergezinnen. Van het kindertal gaat geen effect uit op de toetsscore. Leerlingen waarvan de ouders geen vragenlijst hebben ingevuld, scoren lager op beide toetsen.

Door toevoeging van *cirkel vier* worden ook de autoritatieve opvoedingsstijl en het aantal contacturen in het model opgenomen. De opvoedingsstijl heeft een tamelijk grote invloed op de toetsscores. Wanneer de overige kenmerken constant worden gehouden, dan komt een afwijking van één standaarddeviatie in de autoritatieve opvoedingsstijl neer op een verschil van 0,08 standaarddeviatie op de beide toetsen. Met de opname van de kenmerken van cirkel vier zijn de schattingen van de vaste parameters van de structurele kenmerken iets kleiner geworden. De opvoedingsstijl en het aantal contacturen verklaren dus voor een deel de invloed van de structurele leerling- en gezinskenmerken.

Vervolgens voeg ik de kenmerken uit *cirkel drie* toe. Praten met ouders en huiswerkcontrole vertonen beide samenhang met de toetsscores. Niet alleen het verband tussen huiswerkcontrole en prestaties, maar ook die tussen praten met ouders en prestaties is negatief.^{xix} Voor de relatie tussen deze kenmerken van onderwijsondersteunend gedrag en de toetsscores geldt dat het uiterst moeilijk is om na te gaan wat oorzaak en gevolg is. Bij de interpretatie van deze effecten is dus voorzichtigheid geboden. Door de extra variabelen veranderen de schattingen van de vaste parameters van de kenmerken sekse en etniciteit sterk. Jongens en allochtone leerlingen blijken namelijk minder vaak te praten met hun ouders en worden vaker aangespoord tot of gecontroleerd op het maken van huiswerk.

Alle algemene leerlingkenmerken (*cirkel twee*) blijken een effect te hebben op de toetsscores. Positief voor de prestaties op beide toetsen is het effect van het zelfbeeld. Voor de prestaties op de toets Nederlands heeft het tijd besteden aan lezen en hobby's een positief effect. Negatief voor de prestaties op beide toetsen zijn de effecten van het hebben van een bijbaan en het op straat rondhangen met vrienden. Uitgaan heeft een negatief effect op de prestaties wiskunde. Opnieuw leidt de toevoeging van een cirkel tot veranderingen van de schattingen van de vaste parameters in de overige cirkels. Het aantal contacturen blijkt bijvoorbeeld niet langer van invloed te zijn op de toetsscore. Dit wordt veroorzaakt door de toevoe-

ging van het leefpatroon. Leerlingen die vaak thuis zijn met volwassenen, hangen minder op straat rond met vrienden en besteden meer tijd aan lezen en hobby's. Na de toevoeging van deze cirkel neemt voor Nederlands de schatting van de vaste parameter van het kenmerk sekse weer toe. Dit hangt samen met het verschil tussen jongens en meisjes in algemene kenmerken. Jongens hebben namelijk gemiddeld genomen een positiever zelfbeeld dan meisjes, hangen meer op straat rond, besteden meer tijd aan een bijbaantje en minder tijd aan lezen en hobby's.

Vervolgens zijn in Bijlage 3A en 3B de schoolse leerlingkenmerken (*cirkel één*) opgenomen in de analyse. De kenmerken uit deze cirkel zijn in twee stappen ingevoerd, eerst zonder en vervolgens met opname van de scores op de prestatietoetsen uit het eerste leerjaar. Op deze manier kan de invloed van de kenmerken uit cirkel vijf tot en met twee worden nagegaan in een model waarin geen rekening wordt gehouden met de eerdere prestaties, maar wel met de overige schoolse leerlingkenmerken. Dit model wordt aangegeven als cirkel 1'.

In het model met alle kenmerken uit cirkel één blijkt uiteraard dat de eerdere prestaties een grote invloed hebben op de toetsscores. Er zijn echter ook nog talrijke andere kenmerken die een effect hebben, namelijk non-verbale intelligentie, aantal dagen huiswerk, mening over docenten en regels (cirkel één), zelfbeeld, op straat met vrienden (cirkel twee), huiswerkcontrole (cirkel drie), autoritatieve opvoedingsstijl (cirkel vier), sekse, opleidingsniveau moeder en etniciteit (cirkel vijf). Een effect op de toetsscore Nederlands is er voor mening over klasgenoten (cirkel één), bijbaan, lezen en hobby's (cirkel twee) en opleidingsniveau vader (cirkel vijf). Een effect op de toetsscore wiskunde is er voor uitgaan (cirkel twee), praten met ouders (cirkel drie) en het aantal ouders (cirkel vijf). Wanneer alle cirkels zijn opgenomen, dan is er geen effect meer van de positie in de kinderrij.

Variantiecomponenten en modelfit

In Tabel 3.4 vergelijk ik de variantiecomponenten en de modelfit van de verschillende modellen. Allereerst geef ik het model weer waarin geen rekening wordt gehouden met kenmerken van scholen of leerlingen, het zogenoemde lege model. De *variantiecomponenten* bedragen bij dit model voor Nederlands op leerlingniveau 0,522, op klasniveau 0,229 en op schoolniveau 0,312. De intra-klasse correlatiecoëfficiënt voor verschillen tussen klassen en scholen is $(0,229 + 0,312) / (0,522 + 0,229 + 0,312) = 0,51$.

In landen waar de verschillen in leerlingpopulatie tussen scholen groot zijn, zoals Nederland, kan de variantie in bruto schoolprestaties zo hoog zijn (OECD, 1998). Deze bruto variantie maakt niet duidelijk of scholen verschillen in kwaliteit. Om scholen met elkaar te vergelijken, moet eerst worden gecontroleerd voor

kenmerken op individueel niveau. Ter illustratie, wanneer er per schooltype wordt geanalyseerd blijkt de relatieve tussenschoolse variantie 5 tot 10 procent (Van der Werf, Kuyper & Lubbers, 1999).

Voor verschillen tussen alleen scholen kan de *intra-klasse correlatiecoëfficiënt* als volgt worden berekend: $0,312 / (0,229 + 0,312) = 0,58$. Dit betekent dat meer dan de helft van de variantie tussen klassen en scholen tussenschools is. Voor wiskunde is de variantie tussen klassen en scholen 65 procent en tussen klassen binnen scholen 59 procent. Wanneer er geen rekening wordt gehouden met de leerlingenpopulatie van klassen en scholen dan zijn de verschillen tussen klassen en scholen in het Nederlandse voortgezet onderwijs groot, wat een gevolg is van de aanwezigheid van veel verschillende schooltypen. Deze verschillen tussen schooltypen zitten in de klas- en schoolvariantie.

Tabel 3.4: Multiniveau analyse. Variantie- en covariantiecomponenten en modelfit

	VARIANTIE NEDERLANDS	VARIANTIE WISKUNDE	COVARIANTIE	DEVIANCE)
LEEG MODEL	PAR. (S.E.)	PAR. (S.E.)	PAR. (S.E.)		
Leerlingniveau	0,522 (0,009)	0,396 (0,008)	0,138 (0,007)	26.925,4	
Klasniveau	0,229 (0,021)	0,299 (0,026)	0,236 (0,021)		
Schoolniveau	0,312 (0,048)	0,433 (0,065)	0,342 (0,053)		
CIRKEL 5					
Leerlingniveau	0,507 (0,009)	0,377 (0,007)	0,140 (0,006)	26.329,2	596,2
Klasniveau	0,206 (0,019)	0,276 (0,024)	0,213 (0,019)		(df=20)
Schoolniveau	0,258 (0,040)	0,394 (0,059)	0,294 (0,046)		p<0,001
CIRKEL 4					
Leerlingniveau	0,500 (0,009)	0,371 (0,007)	0,134 (0,006)	26.179,2	150,0
Klasniveau	0,195 (0,018)	0,265 (0,023)	0,203 (0,018)		(df=4)
Schoolniveau	0,239 (0,037)	0,374 (0,056)	0,275 (0,044)		p<0,001
CIRKEL 3					
Leerlingniveau	0,493 (0,009)	0,367 (0,007)	0,129 (0,006)	26.025,9	153,3
Klasniveau	0,188 (0,018)	0,262 (0,023)	0,197 (0,018)		(df=4)
Schoolniveau	0,230 (0,036)	0,369 (0,056)	0,268 (0,043)		p<0,001
CIRKEL 2					
Leerlingniveau	0,475 (0,009)	0,355 (0,007)	0,116 (0,006)	25.668,5	357,4
Klasniveau	0,173 (0,017)	0,247 (0,022)	0,183 (0,017)		(df=10)
Schoolniveau	0,212 (0,034)	0,344 (0,052)	0,246 (0,039)		p<0,001

CIRKEL 1'

Leerlingniveau	0,469 (0,009)	0,335 (0,006)	0,108 (0,006)	25.213,7	454,8
Klasniveau	0,151 (0,015)	0,198 (0,018)	0,150 (0,014)		(df=10)
Schoolniveau	0,186 (0,030)	0,289 (0,044)	0,207 (0,034)		p<0,001

CIRKEL 1

Leerlingniveau	0,421 (0,008)	0,271 (0,005)	0,065 (0,005)	22.992,1	2.221,6
Klasniveau	0,052 (0,007)	0,050 (0,006)	0,029 (0,005)		(df=6)
Schoolniveau	0,062 (0,011)	0,094 (0,015)	0,050 (0,010)		p<0,001

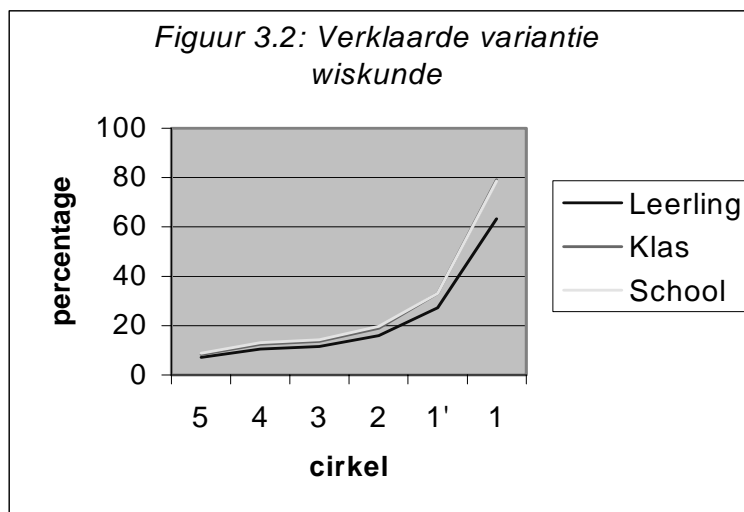
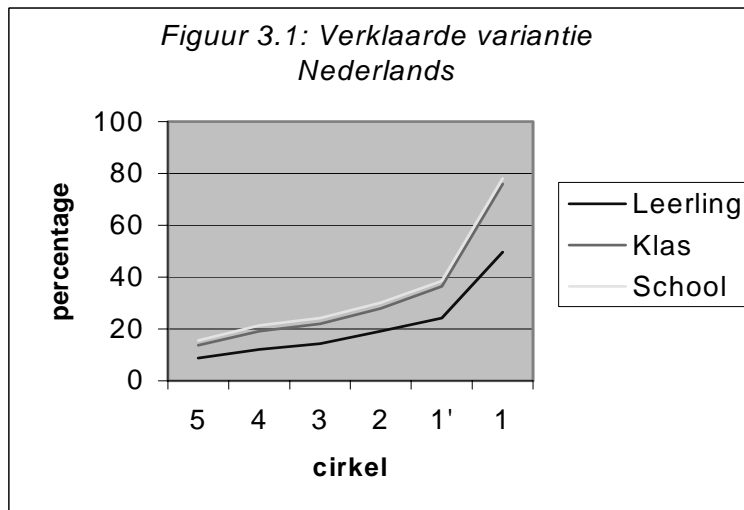
SPAARZAAM MODEL

Leerlingniveau	0,421 (0,008)	0,271 (0,005)	0,065 (0,005)	23.019,5	27,4
Klasniveau	0,052 (0,007)	0,050 (0,006)	0,029 (0,005)		(df=22)
Schoolniveau	0,063 (0,011)	0,094 (0,015)	0,049 (0,010)		p>0,20

In Tabel 3.4 worden ook de *covariantiecomponenten* tussen Nederlands en wiskunde vermeld. Met de schattingen van de varianties en covariantie kan op elk niveau de *correlatie tussen Nederlands en wiskunde* worden berekend. In het lege model is de correlatie 0,93 op school-, 0,90 op klas- en 0,30 op leerlingniveau. Voor het schoolniveau kan deze correlatie als volgt worden berekend: $0,342 / \sqrt{(0,312 * 0,433)} = 0,93$. Naarmate de prestaties van een klas of een school op het ene vak hoger zijn, zijn ook de prestaties op het andere vak hoger. In een klas hoeven leerlingen die hoog scoren op Nederlands niet per se hoog te scoren op wiskunde, en omgekeerd. De samenhang tussen beide toetsen is dus zwak op leerlingniveau (binnen klassen binnen scholen).

In het volgende model zijn de structurele kenmerken toegevoegd. De kenmerken uit deze cirkel verklaren variantie op zowel leerling-, klas- als schoolniveau: 8,7, 13,7 en 15,6 procent voor Nederlands en 7,2, 8,4 en 8,7 procent voor wiskunde. Voor Nederlands kan dit op *leerlingniveau* als volgt worden berekend: $1 - (0,507 + 0,206 + 0,259) / (0,522 + 0,229 + 0,312) = 0,087$. Op *klasniveau* luidt de formule: $1 - (0,507/20 + 0,206 + 0,259) / (0,522/20 + 0,229 + 0,312) = 0,137$ en op *schoolniveau* ziet de vergelijking er als volgt uit: $1 - (0,507/60 + 0,206/3 + 0,259) / (0,522/60 + 0,229/3 + 0,312) = 0,156$. Voor Nederlands wordt op leerlingniveau door de toevoeging van de vierde cirkel 3,4 procent, de derde cirkel 2,2 procent, de tweede cirkel 4,8 procent en de complete eerste cirkel 30,6 procent extra variantie verklaard. Voor wiskunde bedragen deze percentages op leerlingniveau 3,3 voor de vierde, 1,0 voor de derde, 4,6 voor de tweede en 47,1 procent voor de complete eerste cirkel. De extra verklaarde variantie die de eerste cirkel toevoegt komt voornamelijk door de opname van de eerdere prestaties, zo blijkt uit een vergelijking van het laatste en het voorlaatste model. Wanneer ik alleen de mening over klasgenoten, de mening over docenten en regels, het uitstellen van huiswerk,

het aantal dagen dat huiswerk wordt gemaakt en de non-verbale intelligentie opneem (cirkel 1'), neemt de verklaarde variantie op leerlingniveau voor Nederlands en wiskunde met 5,1 en 11,0 procent toe. In totaal wordt op leerlingniveau voor Nederlands 49,7 procent en voor wiskunde 63,2 procent van de variantie verklaard door de vijf cirkels. Op klas- en schoolniveau bedragen de percentages 76,2 en 78,3 voor Nederlands en 79,0 en 78,6 voor wiskunde. Voor de verklaring van verschillen in prestaties tussen leerlingen, klassen en scholen zijn op individueel niveau belangrijke effecten te vinden. In Figuur 3.1 en 3.2 is de verklaarde variantie voor beide vakken grafisch samengevat.



Op leerlingniveau binden de *structurele kenmerken*, cirkel vijf, dus 8,7 en 7,2 procent van de variantie in de toetsscores. De *culturele kenmerken*, de overige cirkels, binden 41,0 en 56,0 procent van de variantie. Deze variantie valt op te splitsen in twee delen, te weten 25,5 en 36,1 procent door de opname van de eerdere

prestaties en 15,5 en 19,9 procent door de opname van de veertien overige culturele kenmerken. In beide grafieken valt te zien dat vooral de opname van de eerdere prestaties tot een sterke stijging van de verklaarde variantie leidt.

Op zowel klas- als schoolniveau binden de structurele kenmerken ongeveer 15 en 8 procent van de variantie in de toetsscores en de culturele kenmerken 62 en 71 procent. Deze variantie wordt voor tweederde door de eerdere prestaties en voor eenderde door de overige culturele kenmerken verklaard, zie Figuur 3.1 en 3.2. De lijnen voor de verklaarde variantie op klas- en schoolniveau zijn niet of nauwelijks van elkaar te onderscheiden. Op beide niveau is de variantie steeds vrijwel gelijk.

In het laatste model in Tabel 3.4 bedragen voor Nederlands de variantie-componenten op leerlingniveau 0,421, op klasniveau 0,052 en op schoolniveau 0,062. De residuele intra-klasse correlatiecoëfficiënt tussen klassen en scholen is 0,21 en tussen scholen 0,54. Voor wiskunde bedraagt de variantie tussen klassen en scholen 35 en tussen scholen 65 procent. Het is gewoon dat de variantie tussen klassen en scholen voor wiskunde hoger is dan voor Nederlands. Wiskunde leren leerlingen vooral op school, Nederlands leren leerlingen zowel op als buiten school. Voor hun prestaties op wiskunde maakt het daarom uit wat leerlingen op school aangeboden krijgen aan lesstof. Voor Nederlands is dat wat minder belangrijk. Varianties tussen klassen en scholen van 21 en 35 procent zijn vrij hoog. In hoofdstuk 4 zal blijken dat dit percentage afneemt door toevoeging van het schooltype waarop een leerling zit. In het volledige model is de correlatie 0,65 op school-, 0,58 op klas- en 0,19 op leerlingniveau.

Door de opname van extra cirkels blijkt de zogeheten deviance steeds af te nemen, zie Tabel 3.4. De afname in deviance is in alle gevallen significant. De opname van een extra cirkel leidt dus steeds tot een verbetering van de verklaring van de beide toetsscores.

Vergelijkbare of verschillende effecten

In Tabel 3.5 geef ik een spaarzaam model weer. Bij het opstellen van dit model heb ik de niet-significante variabelen uit de analyse verwijderd. Vervolgens ben ik voor de overgebleven variabelen nagegaan of ze een *vergelijkbaar of verschillend effect* hebben op Nederlands en wiskunde. In Tabel 3.5 geef ik het verschil tussen elk paar van effecten weer. Het contrast is berekend door van de parameter voor Nederlands de parameter voor wiskunde af te trekken. Het contrast is dus groter dan nul als het effect op Nederlands groter is dan het effect op wiskunde. Omgekeerd is het contrast kleiner dan nul als het effect op wiskunde groter is dan het effect op Nederlands. Het verschil tussen deze parameters kan worden getoetst met een chi-kwadraattoets met één vrijheidsgraad. De effecten die significant van elkaar verschillen zijn vet weergegeven. Hierbij gaat het om de volgende effecten:

aantal ouders, sekse (cirkel vijf), praten met ouders, huiswerkcontrole (cirkel drie), lezen en hobby's, uitgaan (cirkel twee), mening over klasgenoten, non-verbale intelligentie en de drie prestatietoetsen uit leerjaar één (cirkel één). De andere variabelen hebben een vergelijkbaar effect op beide toetsen. Voor deze effecten kan daarom één parameter worden geschat.

*Tabel 3.5: Multiniveau analyse. Spaarzaam model met contrasten (chi-kwadraat verdeeld, $df=1$) en **verschillende** of vergelijkbare effecten*

	CONTRAST	O ²	NEDERLANDS	WISKUNDE
Etniciteit (1=allochtoon)	-0,04	0,59	-0,082 (0,027) **	-0,082 (0,027) **
Opleidingsniveau vader	0,01	0,32	0,024 (0,007) ***	0,024 (0,007) ***
Opleidingsniveau moeder	0,01	0,96	0,021 (0,007) **	0,021 (0,007) **
Aantal ouders (1=twee)	-0,06	4,74	0,001 (0,024)	0,057 (0,021) **
Sekse (1=jongen)	-0,31	177,30	-0,180 (0,020) ***	0,131 (0,017) ***
Géén oudervragenlijst	0,05	2,22	-0,027 (0,019)	-0,027 (0,019)
Autoritatieve opvoedingsstijl	0,01	0,99	0,033 (0,006) ***	0,033 (0,006) ***
Praten met ouders	0,03	6,74	0,006 (0,009)	-0,024 (0,007) ***
Huiswerkcontrole	-0,03	7,98	-0,058 (0,009) ***	-0,026 (0,007) ***
Op straat met vrienden	0,00	0,00	-0,036 (0,007) ***	-0,036 (0,007) ***
Lezen en hobby's	0,05	12,97	0,054 (0,009) ***	0,007 (0,008)
Bijbaan	-0,01	1,61	-0,018 (0,006) **	-0,018 (0,006) **
Uitgaan	0,03	8,97	0,008 (0,009)	-0,024 (0,008) **
Zelfbeeld	0,01	1,13	0,046 (0,007) ***	0,046 (0,007) ***
Mening over klasgenoten	0,04	9,79	0,041 (0,009) ***	0,004 (0,008)
Mening over docenten/regels	0,01	0,94	0,033 (0,006) ***	0,033 (0,006) ***
Aantal dagen huiswerk	-0,01	0,38	0,049 (0,007) ***	0,049 (0,007) ***
Non-verbale intelligentie	-0,08	45,04	0,027 (0,010) **	0,106 (0,009) ***
Prestatie taal	0,12	43,05	0,200 (0,015) ***	0,078 (0,011) ***
Prestatie infoverwerking	0,09	32,96	0,238 (0,013) ***	0,147 (0,011) ***
Prestatie rekenen	-0,30	296,94	0,086 (0,013) ***	0,382 (0,014) ***
Sekse * Taal			-0,041 (0,017) *	
Lezen en hobby's * Taal			0,031 (0,007) ***	
Sekse * Rekenen				-0,054 (0,015) ***
Constante			0,074 (0,034)	-0,148 (0,035)

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Van een groot aantal van de effecten is het theoretisch plausibel dat ze verschillend zijn naar vak. Dit geldt bijvoorbeeld voor het sekse-effect, want conform de verwachting presteren jongens beter op wiskunde en meisjes beter op Nederlands. Ook was de verwachting dat het effect van lezen en hobby's zou verschillen naar vak. Uit Tabel 5.3 blijkt dat er wel een effect is van dit kenmerk op Nederlands en niet op wiskunde. De vakspecifieke effecten voor de non-verbale intelligentie en de eerdere prestaties zijn ook aannemelijk. De grootte van de effecten verschilt naar vak doordat enerzijds taal en informatieverwerking meer samenhangen met Nederlands en anderzijds non-verbale intelligentie (redeneren en abstraheren) en rekenen meer gerelateerd zijn aan wiskunde. Voor de overige kenmerken is niet duidelijk waarom er verschillen zijn naar vak.

Voor dit spaarzame model wil ik tot slot uiteenzetten hoe belangrijk de verschillende effecten zijn.^{xx} Onder constant houding van de overige kenmerken heeft de *seks* een sterk effect op de prestaties van leerlingen. Meisjes blijken gemiddeld genomen beduidend beter te scoren op Nederlands en jongens op wiskunde. Wanneer deze effecten worden vergeleken met het gangbare verschil van een halve standaarddeviatie tussen schooltypen (zie Tabel 3.2, het verschil tussen de toetsscores van bijvoorbeeld havo en vwo is 0,49 op Nederlands en 0,51 op wiskunde), dan kan worden geconcludeerd dat in een gemiddelde havo-klas vier op de tien jongens voor Nederlands presteren op mavo-niveau en twee of drie voor wiskunde op vwo-niveau. Voor meisjes geldt het omgekeerde, in een gemiddelde havo-klas presteren vier op de tien meisjes voor Nederlands op vwo-niveau en twee of drie voor wiskunde op mavo-niveau.

De effecten van *etniciteit en opleidingsniveau van ouders* wil ik verduidelijken door ze in samenhang te bekijken. Stel, een allochtone leerling heeft ouders met niet meer dan basisonderwijs. Wanneer deze leerling wordt vergeleken met een autochtone leerling met ouders met een hogere beroepsopleiding, dan is het verschil tussen deze leerlingen voor beide vakken bijna een kwart standaarddeviatie (0,082 voor het effect van etniciteit plus een verschil van ongeveer drie standaarddeviaties in het effect van beide opleidingsniveaus, dus: $3 * 0,024 + 3 * 0,021$), waarbij in het spaarzame model gecontroleerd is voor alle relevante kenmerken uit cirkel vijf. Als deze effecten in samenhang worden bekeken is dus sprake van een sterk effect van de sociaal-economische status en etniciteit op prestaties van leerlingen. Net als in het merendeel van de Nederlandse onderzoeken naar verschillen in prestaties zijn er effecten van zowel de opleiding van de moeder en de vader als van de etniciteit. Het verschil tussen allochtonen en autochtonen valt niet weg, wanneer er rekening wordt gehouden met de kenmerken van de sociaal-economische status. Etniciteit

heeft dus een extra effect op prestaties en dan niet alleen op Nederlands maar ook op wiskunde. Van het opgroeien in een *tweeoudergezin* gaat een positief effect uit op de wiskundeprestaties. Het effect van het aantal ouders is 0,057. In het spaarzame model is er geen effect meer van het *ontbreken van een oudervragenlijst*. Na controle voor de kenmerken uit cirkel één, verschillen leerlingen waarvan de ouders geen vragenlijst hebben ingevuld niet systematisch van de overige leerlingen.

Behalve effecten van structurele kenmerken zijn er in het spaarzame model ook effecten van culturele kenmerken. Van de kenmerken uit cirkel drie en vier hebben de *huiswerkcontrole* en de *autoritatieve opvoedingsstijl* een behoorlijke invloed op de prestaties van leerlingen. Stel een leerling wordt niet door zijn of haar ouders aangespoord tot en gecontroleerd op het maken van huiswerk en verder wordt de helft van de beslissingen door de leerling genomen na overleg met de ouders. Wanneer deze leerling wordt vergeleken met een leerling die geregeld te maken heeft met huiswerkcontrole, bijvoorbeeld een score van 6 op de betreffende schaal, en die niet autoritatief wordt opgevoed, dan is het verschil tussen deze leerlingen op huiswerkcontrole vier standaarddeviaties en op de autoritatieve opvoedingsstijl drie standaarddeviaties. Voor Nederlands resulteert dit in een verschil van eenderde standaarddeviatie en voor wiskunde in een verschil van eenvijfde standaarddeviatie. *Praten met ouders* hangt alleen samen met de prestaties van leerlingen op wiskunde. Opvallend genoeg is deze samenhang tegengesteld aan de verwachting. Van leerlingen die meer met hun ouders over school en andere onderwerpen praten, zijn de prestaties op wiskunde slechter. Deze relatie is overigens zwak.

Van de kenmerken uit cirkel twee heeft het *zelfbeeld* van leerlingen het sterkste effect. Dit effect is net zo groot als het gezamenlijke effect van het opleidingsniveau van beide ouders. Om de vier kenmerken van het *leefpatroon* van leerlingen te beschouwen, wil ik opnieuw twee leerlingen als voorbeeld nemen. De ene leerling besteedt van zijn of haar vrije tijd 15 procent aan uitgaan, 15 procent aan een bijbaan, 15 procent op straat met vrienden en besteedt geen tijd aan lezen of hobby's. Voor de andere leerling geldt dat hij of zij niet uitgaat, geen bijbaan heeft en niet op straat met vrienden rondhangt. Aan lezen en hobby's gaat voor deze leerling echter 20 procent van de vrije tijd op. Het verschil tussen deze leerlingen is twee standaarddeviaties op uitgaan, bijbaan en op straat met vrienden en drie standaarddeviaties op lezen en hobby's, wat resulteert in een verschil van een kwart standaarddeviatie op Nederlands en eenvijfde standaarddeviatie op wiskunde.

De grootste effecten op de prestaties komen natuurlijk van de effecten van de *eerdere prestaties*. Wanneer een leerling bijvoorbeeld op alledrie de toetsen één standaarddeviatie boven het gemiddelde scoort, presteert een leerling twee jaar later op havo-niveau. Bij twee standaarddeviaties boven het gemiddelde scoort die

leerling in leerjaar drie op vwo-niveau. Een verschil in één standaarddeviatie resulteert in een verschil van ongeveer één schooltype. De eerdere prestaties zijn daarmee een goede voorspeller van de latere prestaties van leerlingen. Voor de andere kenmerken uit cirkel één geldt dat het aantal dagen huiswerk een behoorlijke invloed heeft. Wanneer een leerling in plaats van vijf maar twee dagen in de week huiswerk maakt resulteert dat in afname van de prestaties op beide toetsen met een tiende standaarddeviatie. De mening over docenten en regels heeft een effect op beide toetsen dat vergelijkbaar is met dat van de autoritatieve opvoedingsstijl. Hiermee vergelijkbaar zijn de effecten van de mening over klasgenoten en de non-verbale intelligentie op de prestaties Nederlands. Het effect van de non-verbale intelligentie op wiskundeprestaties is bijna vier keer zo groot.

Interacties

Behalve naar hoofdeffecten worden in Tabel 3.5 ook interactie-effecten weergegeven.^{xxi} De analyses met interactie-effecten zijn eerst per afhankelijke variabele uitgevoerd. Uit de analyses komen drie interacties naar voren die een significante bijdrage leveren aan de verklaring van de toetsscore Nederlands, te weten sekse met de prestatie op de taaltoets in leerjaar één, lezen en hobby's met wederom de prestatie op de taaltoets in leerjaar één, en sekse met de rekenprestaties in het eerste leerjaar. De relevante termen voor deze interacties zijn: $0,074 - 0,180*(\text{sekse}) + 0,200*(\text{taal}) - 0,041*(\text{sekse} * \text{taal})$ en $0,074 + 0,054*(\text{lezen en hobby's}) + 0,200*(\text{taal}) + 0,031*(\text{lezen en hobby's} * \text{taal})$ en $-0,148 + 0,131*(\text{sekse}) + 0,382*(\text{rekenen}) - 0,054*(\text{sekse} * \text{rekenen})$. Voor leerlingen die in leerjaar één goed presteerden op taal geldt dat ze twee jaar later extra goed presteren op Nederlands als ze een meisje zijn of naarmate ze meer tijd besteden aan lezen en hobby's. Voor leerlingen die in het eerste leerjaar goed presteerden op rekenen geldt dat ze twee jaar later extra goed presteren op wiskunde als ze wederom een meisje zijn. In vergelijking met het model met alle hoofdeffecten neemt de deviance van het model met interactietermen significant af. De afname is namelijk 25,4 (df=3, $p<0,001$). De variantiecomponenten veranderen nauwelijks door de opname van deze drie effecten.

3.5 Besluit

Talrijke leerling- en gezinskenmerken hebben invloed op de prestaties Nederlands en wiskunde van leerlingen in het derde leerjaar van het voortgezet onderwijs. Het niveau waarop leerlingen presteren op Nederlands en wiskunde wordt bepaald door zowel structurele als culturele kenmerken. Om na te gaan welke leerlingen een voorsprong of achterstand hebben op Nederlands en wiskunde is het, zelfs als

er rekening wordt gehouden met de eerdere prestaties van leerlingen, van belang om te weten wat de *opleiding van hun ouders* is en of ze *autochtoon of allochtoon* zijn. Wanneer het effect van deze kenmerken in samenhang wordt bekeken, dan is er sprake van een sterk effect. Van de structurele kenmerken is het onderscheid tussen jongens en meisjes het meest belangrijk voor de verklaring van de prestaties. Onder constant houding van de andere kenmerken, scoren jongens 0,19 standaarddeviatie lager op Nederlands dan meisjes, maar op wiskunde juist 0,13 standaarddeviatie hoger. Van het opgroeien in een *tweeoudergezin* gaat een positief effect uit op de wiskundeprestaties.

Naast effecten van structurele kenmerken zijn er ook effecten van culturele kenmerken. Van de interactie tussen ouders en leerlingen hebben de *huiswerkcontrole* en de *autoritatieve opvoedingsstijl* een behoorlijke invloed op de prestaties van leerlingen. Gedrag van ouders blijkt dus eveneens van belang voor de prestaties van leerlingen. Een kenmerk van de autoritatieve opvoedingsstijl is dat ouders kinderen zelfverantwoordelijkheid laten dragen zonder dat ze zichzelf wegcijferen. Ouders en kinderen overleggen daarbij samen over belangrijke keuzen, maar het kind is uiteindelijk verantwoordelijk. Onder constant houding van de overige kenmerken uit de vijf cirkels, komt een afwijking van één standaarddeviatie in de autoritatieve opvoedingsstijl neer op een verschil van 0,04 standaarddeviatie op de toets.

Het blijkt dat de samenhang tussen de huiswerkcontrole door ouders, een vorm van 'onderwijsondersteunend' gedrag, en de prestaties van hun kinderen negatief is. Wanneer de overige kenmerken constant worden gehouden, dan komt een afwijking van één standaarddeviatie in de huiswerkcontrole neer op een verschil van 0,06 en 0,03 op de toets. Dit negatieve effect van huiswerkcontrole herhaaldelijk in onderwijsonderzoek wordt gevonden. Moeten ouders deze controle nalaten? Of zou het zo kunnen zijn dat ouders meer gaan controleren naarmate de schoolprestaties van hun kind lager zijn of lager worden? Met de gegevens uit VOCL'93 is deze vraag niet goed te beantwoorden. Voor de beantwoording van deze vraag zou er een onderzoeksdesign moeten worden gekozen, waarbij het onderwijsondersteunend gedrag van ouders en de prestaties van leerlingen herhaaldelijk worden gemeten, bijvoorbeeld elke maand.

Van de algemene kenmerken van leerlingen heeft het *zelfbeeld* van leerlingen een sterk, positief effect. Dit effect is net zo groot als het gezamenlijke effect van het opleidingsniveau van beide ouders. Ook het *leefpatroon* heeft een sterke invloed op de prestaties van leerlingen. De grootste effecten op de prestaties komen natuurlijk van de effecten van schoolse leerlingkenmerken en dan met name van de *eerdere prestaties*. Voor de andere schoolse leerlingkenmerken geldt dat het aantal dagen huiswerk een behoorlijke invloed heeft.

Door alle vijf cirkels gezamenlijk wordt op leerlingniveau voor Nederlands bijna 50 procent en voor wiskunde meer dan 60 procent van de variantie gebonden. Op klas- en schoolniveau wordt voor beide toetsen meer dan driekwart van de variantie verklaard (Figuur 3.1 en 3.2).

Met de structuur van concentrische cirkels kunnen de invloeden van de leerling- en gezinskenmerken goed worden weergegeven. De resultaten stemmen overeen met de literatuur. Met uitzondering van het effect van praten met ouders op prestaties wiskunde, correspondeert de richting van de overige effecten steeds met eerder onderzoek. Uniek aan een groot databestand als VOCL'93 is dat veel verschillende leerling- en gezinskenmerken tegelijkertijd kunnen worden bekeken. Wat uit verschillende onderzoeken apart bekend is, heb ik nu in samenhang kunnen onderzoeken.

Op grond van theoretische overwegingen is de structuur van concentrische cirkels opgesteld. Bij het opstellen van deze structuur was één van de doelen dat de effecten van structurele kenmerken op de afhankelijke variabelen verklaard kunnen worden door culturele kenmerken. Uit dit hoofdstuk blijkt dat met deze structuur de verschillende invloeden goed kunnen worden geordend en dat ook inzicht kan worden gegeven in de (causale) relaties tussen de kenmerken onderling. Zo blijken de effecten van kenmerken als de opleiding van de moeder of de etniciteit voor een belangrijk deel te verklaren door intermediaire variabelen. Beide effecten worden voor ongeveer drievijfde verklaard door culturele kenmerken. Overigens kunnen niet alle effecten van structurele kenmerken evengoed worden verklaard. De sterkte van het effect van de opleiding van de vader neemt niet of nauwelijks af door de opname van culturele kenmerken. Om dat effect te verklaren door intermediaire variabelen moet dus aan andere culturele kenmerken worden gedacht.

Bij het opstellen van deze structuur zijn twee aannames gedaan over de volgorde van de culturele kenmerken. De eerste aanname was dat de prestaties van leerlingen meer worden verklaard door kenmerken van leerlingen dan van gezinnen. Deze aanname is plausibel. Uit het literatuuroverzicht komen veel meer kenmerken van leerlingen dan van de gezinsinteractie naar voren. Ook blijkt dat de verklaarde variantie veel sterker toeneemt door de opname van leerlingkenmerken (cirkel 1' en 2) dan door gezinskenmerken (cirkel 3 en 4), zie Figuur 3.1 en 3.2 of de laatste kolom van Tabel 3.4.

De tweede aanname is dat de prestaties van leerlingen meer door schoolse dan door algemene kenmerken worden verklaard. Voor leerlingkenmerken lijkt deze aanname plausibel, maar voor gezinskenmerken geldt dat niet. De verklaarde variantie verschilt niet tussen de twee cirkels voor de interactie in het gezin.

Hieruit kan worden geconcludeerd dat de structuur van concentrische cirkels mogelijk kan worden vereenvoudigd door de schoolse en algemene gezinskenmerken samen te voegen. Voor deze conclusie pleit ook dat de sterkte van de effecten van algemene gezinskenmerken op prestaties van leerlingen nauwelijks afneemt door de opname van schoolse gezinskenmerken, zie Bijlage 3A en 3B.

Een belangrijk pluspunt van de analyses uit dit hoofdstuk is dat met multivariate modellen kan worden nagegaan of variabelen een vergelijkbaar of verschillend effect hebben op de afhankelijke variabelen. Ook is het een voordeel dat nu de correlatie tussen de afhankelijke variabelen op verschillende niveaus kan worden bepaald. Het blijkt dat klassen en scholen waar de prestaties gemiddeld op het ene vak hoog zijn ook de klassen en scholen zijn waar de prestaties op het andere vak hoog zijn, wat ook logisch is aangezien leerlingen van ivbo tot en met vwo in de analyse zijn opgenomen. Voor leerlingen binnen klassen binnen scholen geldt echter dat de samenhang tussen prestaties op Nederlands en wiskunde zwak is. Ook geldt dat leerlingen met een ontbrekende waarde op een van de afhankelijke variabelen in de analyse kunnen worden meegenomen. Het is dus niet nodig dat leerlingen op alle afhankelijke variabelen een score hebben. Een multivariaat multiniveau model heeft dus belangrijke voordelen boven de modellen met analyses voor afzonderlijke afhankelijke variabelen.

Eindnoten

ⁱ De structuur van concentrische cirkels heb ik eerder gebruikt voor een analyse met alleen Nederlands als criteriumvariabele (Veenstra & Kuyper, 1998).

ⁱⁱ Succesvolle Marokkaanse en Turkse leerlingen van de tweede generatie blijken vooral ondersteuning te krijgen van oudere broers en zussen. Het oudste kind, die geen gebruik kan maken van dergelijke hulp, doet het dan ook minder goed dan kinderen met een andere positie in de kinderrij. Van de schoolgaande Marokkaanse en Turkse leerlingen volgt het tweede, derde of vierde kind twee maal zo vaak een havo- of vwo-opleiding als het eerste kind (Crul, 1996).

ⁱⁱⁱ Een ander kenmerk dat in deze cirkel zou passen is het *leefpatroon van ouders*. In navolging van Bourdieu (1977) hebben verscheidene onderzoekers, waaronder Ganzeboom (1984), De Graaf (1987), Van der Velden (1991) en Niehof (1997), laten zien dat er een effect uitgaat van het cultureel kapitaal van ouders op de prestaties van hun kinderen. Het leefpatroon van ouders beïnvloedt waarschijnlijk ook het leefpatroon en de inzet van leerlingen. In dit onderzoek is dit kenmerk niet opgenomen, omdat de gegevens hierover aanvankelijk niet beschikbaar waren. In een extra oudervragenlijst is dit kenmerk inmiddels gemeten (Van der Werf & Kuyper, 1998). In toekomstig onderzoek kan dit kenmerk dan ook worden opgenomen in de structuur van concentrische cirkels.

^{iv} Een ander kenmerk dat in deze cirkel zou passen is de *regelcompetentie* van leerlingen (Van Liere, 1990). Een regelcompetente leerling is goed in staat om zich in sociale situaties te verplaatsen

en de regels en verwachtingen die daar gelden te onderkennen en begrijpen. Ook kan een dergelijke leerling zich goed aan deze regels en verwachtingen houden wanneer daar geen onmiddellijke externe beloningen of sancties tegenover staan. In dit onderzoek zijn over het regelcompetente gedrag van leerlingen geen gegevens beschikbaar.

Ook zouden in deze cirkel persoonlijkheidskenmerken (*The Big Five*) van leerlingen passen. Hierbij gaat het om de kenmerken *nauwkeurigheid* (bijvoorbeeld zorgvuldig en systematisch zijn), *openheid* (intellectueel betrokken, nieuwsgierig, creatief en vindingrijk zijn), *extraversie* (spraakzaam en opgewekt zijn), *vriendelijkheid* (hulpvaardig en sympathiek zijn) en *emotionele stabiliteit* (niet snel in paniek zijn). Van Eijck (1999) heeft laten zien dat deze kenmerken samenhangen met iemands opleidingsniveau, ook als er rekening wordt gehouden met de sociaal-economische gezinsachtergrond en de leeftijd van personen.

^v Uit Amerikaans onderzoek blijkt dat ‘zwarte’ verkering een negatief effect op prestaties heeft (Pallas, 1993; Lucas, 1997).

^{vi} In deze cirkel zouden ook de zogeheten *metacognitieve vaardigheden* van leerlingen passen. In meer of mindere mate gebruiken scholieren vaardigheden om voor, tijdens en na het maken van huiswerk of een opdracht het leren te sturen. Deze metacognitieve vaardigheden hebben volgens Wang, Haertel en Walberg (1993) een behoorlijke invloed op de prestaties van leerlingen. Voorbeelden van metacognitieve vaardigheden zijn het maken van een planning, het stellen van tussendoelen en het geregeld nagaan of de tussendoelen zijn behaald. Wanneer leerlingen op een dergelijke manier hun leerprocessen en -activiteiten sturen, dan leidt dat wellicht tot betere prestaties (Simons, 1995). In een lopend onderzoek van het GION (De Jager & Reezigt, 1996) wordt specifiek op dit kenmerk ingegaan.

^{vii} In het gerucht makende boek *The Bell Curve* voeren Herrnstein en Murray (1994) IQ als voorspeller voor armoede, opleidingsgraad, criminaliteit, huwelijk en echtscheiding, et cetera op. De kritiek op dit boek was hevig, zie bijvoorbeeld: Fischer, Hout, Sánchez Jankowski, Lucas, Swidler en Voss (1996).

^{viii} Wanneer naar specifieke groepen wordt gekeken valt het op dat de uitval onder met name Turkse en Marokkaanse en in mindere mate Antilliaanse, Arubaanse, Surinaamse en Molukse ouders onevenredig hoog is (zie ook Brandsma & Van der Werf, 1997). Uit een analyse op de prestaties in leerjaar één, advies voor voortgezet onderwijs en het brugklasniveau blijkt dat kinderen van niet-responderende ouders significant lager hebben gescoord dan de kinderen van responderende ouders (Brandsma, Lugthart & Van der Werf, 1997). Dit verschil geldt voor zowel autochtone als allochtone leerlingen.

^{ix} Van de in 1993-'94 deelnemende leerlingen blijft het CBS de schoolloopbaan volgen. Van deze leerlingen is bijvoorbeeld bekend of ze zijn blijven zitten, doorstromen naar andere vormen van onderwijs en wanneer ze hun eindexamen halen. VOCL'93 is dus een *longitudinaal onderzoek*.

^x Een voorname reden om af te zien van verdere deelname aan VOCL'93 kwam door de *schaalvergroting* in het voortgezet onderwijs. De meeste scholen voor voortgezet onderwijs bestaan tegenwoordig uit verscheidene locaties. Op bijvoorbeeld een brede scholengemeenschap met een heterogeen brugjaar zitten leerlingen in de eerste klas op één locatie. De afname van prestatietoetsen is op dat moment gemakkelijk te organiseren. In de derde klas zijn de leerlingen bij wijze van spreken verdeeld over zes locaties, waardoor het voor een dergelijke school een grote opgave is om de toetsafname te regelen. Een andere belangrijke reden voor deze uitval is dat de *planning van het onderzoek* slecht was. Vlak voor de zomervakantie moesten er toetsen bij

leerlingen worden afgenomen. Precies in een tijd dat op veel scholen de basisvorming werd afgerond, examens werden afgenomen, proefwerkweken op het rooster stonden of excursies werden georganiseerd. Het is naar mijn mening niet vreemd dat sommige directieleden ervan afzagen om leerlingen in die periode een hele dag toetsen te laten maken. Behalve de planning liet ook de *organisatie van de dataverzameling* te wensen over. Door geldgebrek werden er geen testleiders naar scholen gestuurd. Tegen een kleine vergoeding moest het schoolpersoneel zelf de toets afnemen. Aangezien sommige directieleden geen zin hadden om hun eigen personeel met de toetsafname te belasten, zagen er opnieuw scholen van verdere deelname af. Volgens Van der Werf, Lubbers en Kuyper (1999) zijn de resultaten van de toetsingen in leerjaar drie overigens nauwelijks vertekend door de non-respons.

^{xi} In totaal is van 7.493 leerlingen informatie verkregen via de LKS-vragenlijst. De gegevens van deze leerlingen zijn vervolgens gekoppeld met gegevens uit VOCL'93. Voor 101 leerlingen bleek deze koppeling niet mogelijk, omdat ze niet deelnemen aan het cohortonderzoek of een verkeerd koppelnummer hadden gekregen. Voor 178 leerlingen geldt dat ze wel aan VOCL'93 hebben deelgenomen in het schooljaar 1993-'94, maar niet in 1995-'96. Deze leerlingen konden ook niet in dit onderzoek worden meegenomen. Voor deze leerlingen ontbreekt bijvoorbeeld de score op de toetsen Nederlands en wiskunde.

^{xii} Aanvankelijk had ik het idee om van leerlingen niet alleen de brugklas maar ook de derde klas in de structuur van de multiniveau analyse op te nemen. Hiertoe kan gebruik worden gemaakt van cross-classificaties (Goldstein, 1995²; Goldstein & Sammons, 1997). De leerlingen zwermen na de brugklas echter zo veel uiteen dat de gemiddelde grootte van een derde klas in VOCL'93 nog maar vijf leerlingen is. Praktisch gezien lijkt het opnemen van de derde klas van leerlingen dus onaantrekkelijk.

^{xiii} Verwijderd zijn mavo-, havo-, havo/vwo- en vwo-leerlingen die meer dan twee standaarddeviaties beneden het gemiddelde scoorden van leerlingen op hetzelfde schooltype. Deze selectie op de afhankelijke variabele is verricht, omdat er aanwijzingen waren dat de toets Nederlands door sommige leerlingen niet serieus was gemaakt.

^{xiv} Selectiviteit (*selection bias*) kan in VOCL'93 optreden doordat zittenblijven, voortijdig schoolverlaten en verhuizing niet random zijn verdeeld. Bepaalde leerlingen verlaten bijvoorbeeld eerder het onderwijs voortijdig dan anderen. Het is dan allereerst de vraag in hoeverre de onderzochte groep representatief is voor de populatie van leerlingen in leerjaar drie van het voortgezet onderwijs. Vervolgens is het ook de vraag of de parameterschattingen niet zijn vertekend (Berk, 1983; Heckmann, 1979). In vervolgonderzoek zou kunnen worden nagegaan in hoeverre selectiviteit de resultaten vertekent. In dit opzicht moet vooral aan indirecte selectie (Berk, 1983) worden gedacht.

Om deze selectiviteit te onderzoeken heeft Heckmann (1979) een procedure opgesteld om met een probit model na te gaan van welke onderzoekspersonen gegevens ontbreken op een endogene variabele. In dit onderzoek zou dus nagegaan kunnen worden welke leerlingen wel en welke leerlingen niet een score hebben op een van de toetsen. Vervolgens wordt met de voorspelde waarden uit de probit analyse een maat opgesteld die de kans op uitval aangeeft, de zogeheten *hazard rate*. Deze maat wordt dan als onafhankelijke variabele in de analyses op beide toetsscores opgenomen (Berk, 1983). Naast de Heckmann-correctie is er overigens ook een methode van Rubin, gebaseerd op de zogeheten 'propensity score'.

^{xv} Voor het gelijkenschakelen van de toetsen is gebruik gemaakt van het programma OPLM, het One-Parameter Logistic Model (Verhelst, Glas & Verstralen, 1995).

^{xvi} De variabelen van het leefpatroon van leerlingen zijn scheef verdeeld naar rechts. Enkele leerlingen scoorden extreem hoog op één van deze variabelen, bijvoorbeeld 100 procent op de tijdsbesteding aan lezen en hobby's. Om de invloed van deze uitbijters tegen te gaan, is het maximum van deze variabelen gesteld op 35.

^{xvii} Zie Huisman (1999) voor een overzicht van verschillende technieken om ontbrekende waarnemingen te vervangen. Huisman stelt overigens dat het imputeren van gemiddelden doorgaans niet een erg goede manier is.

^{xviii} Aangezien de nominale variabelen niet zijn gestandaardiseerd is de totale variantie in Tabel 3.4 niet precies gelijk aan één.

^{xix} De bivariate correlaties tussen huiswerkcontrole en de toetsscores is zwak negatief, maar tussen praten met ouders en de toetsscores zwak positief.

^{xx} Uit sociologisch oogpunt zijn de voorlaatste modellen uit de Bijlagen 3A en 3B ook van belang. In de modellen met alle kenmerken op individueel niveau met uitzondering van de eerdere prestaties, blijken verscheidene leerling- en gezinskenmerken een sterker effect te hebben dan in het model met de eerdere prestaties. Allochtone leerlingen scoren bijvoorbeeld 0,163 standaarddeviatie lager op Nederlands en 0,141 standaarddeviatie lager op wiskunde dan autochtonen. In het model met eerdere prestaties zijn deze effecten bijna gehalveerd.

^{xxi} Bij gebrek aan duidelijke hypothesen over leerling-leerling interacties heb ik alle mogelijke interacties in een analyse opgenomen (vergelijk Aitkin & Zuzovsky, 1994). Om fouten van de eerste soort tegen te gaan, heb ik via de Bonferroni-procedure de kritieke waarde voor deze tweeweg-interacties verhoogd (Stevens, 1992). De minimale t-waarde voor deze effecten is hierdoor gelijk aan 3,1.

4 | Het klas- en schoolniveau

4.1 Inleiding

Op klas- en schoolniveau wordt, zo blijkt uit het vorige hoofdstuk, meer dan driekwart van de variantie in de beide toetsscores gebonden door kenmerken van leerlingen en gezinnen. In dit hoofdstuk wil ik nagaan of er naast effecten van leerling- en gezinskenmerken ook effecten van docenten en scholen (inclusief contextuele kenmerken) zijn. Deze vraag probeer ik op verschillende manieren te beantwoorden. Allereerst wil ik nog een kenmerk op leerlingniveau toevoegen, namelijk het door een leerling gevolgde *schooltype*. Dit kenmerk heb ik in het vorige hoofdstuk niet opgenomen in de analyses, omdat het schooltype niet alleen een indicatie geeft van het prestatieniveau van een leerling. Het schooltype geeft ook de gelegenheid tot leren weer. Naarmate het schooltype hoger is, krijgen leerlingen moeilijkere leerstof aangeboden, waardoor bijvoorbeeld vwo-leerlingen meer gelegenheid tot leren hebben dan overige leerlingen. Ten tweede wil ik nagaan of er *random effecten* van leerling- en gezinskenmerken zijn. Onafhankelijke variabelen op het leerlingniveau kunnen immers op hogere niveaus differentiële effecten hebben. Met multiniveau analyse kan er in de derde plaats onderscheid worden gemaakt tussen *tussenschoolse* en *binnenschoolse regressie-effecten*. Hierdoor kan worden bekeken of leerling- en gezinskenmerken niet alleen op individueel maar ook op klas- of schoolniveau een effect hebben op prestaties.

Tot slot voeg ik *vaste effecten* van variabelen op *hoger niveau* toe. Hierbij besteed ik aandacht aan docent- en schoolkenmerken. Om de effecten van deze kenmerken te verklaren gebruik ik ook op klas- en schoolniveau de structuur van concentrische cirkels. Op deze niveaus bestaat de structuur uit drie cirkels: (C) contextuele kenmerken; (B) schoolkenmerken; (A) docentkenmerken.

De opzet van dit hoofdstuk is als volgt. In paragraaf 4.2 geef ik een overzicht van de invloed van docent- en schoolkenmerken. Op deze manier krijgt de structuur van concentrische cirkels nadere invulling. In paragraaf 4.3 behandel ik de respons, de operationalisatie en de beschrijvende statistieken van de docent- en schoolkenmerken. In de twee daarop volgende paragrafen ga ik in de op de effecten op klas- en schoolniveau. In paragraaf 4.4 voer ik drie stappen uit. Allereerst breid ik het spaarzame model uit hoofdstuk 3 uit met de variabele schooltype. In

de tweede en derde plaats onderzoek ik in die paragraaf of er random effecten en regressie-effecten tussen klassen en scholen zijn. In paragraaf 4.5 kijk ik naar de vaste effecten van docent- en schoolkenmerken.

In dit hoofdstuk is de algemene onderzoeksvraag: *In hoeverre en waarom zijn kenmerken op klas- en schoolniveau van invloed op de prestaties van leerlingen in het derde leerjaar van het voortgezet onderwijs?* Deze vraag kan worden gesplitst in vier delen:

1. In hoeverre en waarom heeft het schooltype effect op de prestaties van leerlingen in het derde leerjaar van het voortgezet onderwijs, gegeven de invloed van leerling- en gezinskenmerken?
2. Zijn de effecten van leerling- en gezinskenmerken vaste of random effecten?
3. Zijn de effecten van leerling- en gezinskenmerken regressie-effecten binnen of tussen klassen en scholen?
4. In hoeverre en waarom hebben docent- en schoolkenmerken effect op de prestaties van leerlingen in het derde leerjaar van voortgezet onderwijs, gegeven de invloed van leerling- en gezinskenmerken en het schooltype?

4.2 Belang van docent- en schoolkenmerken

In deze paragraaf bespreek ik de docent- en schoolkenmerken die van invloed kunnen zijn op de prestaties van leerlingen. Voor de keuze van deze kenmerken vormt het werk van Creemers (1991; 1994), Scheerens (1989; 1992) en Scheerens en Bosker (1997) het vertrekpunt. Hierbij maak ik een onderscheid tussen contextuele, school- en docentkenmerken.

Hoe ziet een effectieve school eruit? Wat voor activiteiten onderneemt een effectieve docent? Sinds de jaren zeventig worden deze vragen geregeld gesteld. In de loop van de jaren heeft dit tot talrijke opsommingen van kenmerken van effectieve scholen geleid. In hoofdstuk 1 heb ik de kenmerken die Edmonds (1979) als cruciaal beschouwde naar voren gebracht. Op dit werk zijn in latere jaren talrijke variaties ontstaan, waarbij het vijf-factorenmodel van Edmonds meestal met enkele kenmerken wordt uitgebreid (Brophy & Good, 1986; Sammons, Hillman & Mortimore, 1995; Scheerens & Bosker, 1997; Wilson & Corcoran, 1988). Op basis van dergelijke overzichten, kom ik tot een overzicht van verschillende kenmerken die mogelijk van belang zijn voor prestatieverschillen tussen scholen. Allereerst gaat het hierbij om drie contextuele kenmerken: (1) locatiegrootte; (2) leerlingencompositie; (3) denominatie. In de tweede plaats kunnen verscheidene schoolkenmerken van belang zijn: (4) nadruk op basisvaardigheden; (5) professioneel en onderwijskundig leiderschap; (6) eenheid in visie en doelen; (7) orde-

lijk schoolklimaat; (8) bewaking van leervorderingen; (9) samenwerking tussen school en thuis; (10) aard brugperiode; (11) mentoraatsysteem; (12) docent-leerling ratio; (13) toelatingscriteria. De eerste zes schoolkenmerken komen in verscheidene opsommingen voor. De laatste vier worden minder vaak genoemd. Tot de herhaaldelijk genoemde kenmerken van effectieve docenten behoren: (14) gelegenheid tot leren en effectieve leertijd en (15) gestructureerd onderwijs.

Contextuele kenmerken (cirkel C)

De locatiegrootte, de leerlingencompositie en de denominatie zijn alledrie contextuele schoolkenmerken.ⁱ Deze kenmerken komen sterk overeen met de structurele kenmerken op leerlingniveau. Het zijn kenmerken die doorgaans als gegeven worden beschouwd. Voor deze schoolkenmerken kan de schoolleiding of het bestuur niet of nauwelijks beleid voeren.

Wat *locatiegrootte* betreft variëren scholen voor voortgezet onderwijs sterk. In het voortgezet onderwijs was in de jaren tachtig en negentig sprake van schaalvergroting (Peschar & Veenstra, 1999). Het aantal scholen ging van 2.472 in 1980 naar 803 in 1995. Die achteruitgang was het sterkst bij categorale scholen. Zo liep het aantal zelfstandige scholen voor voorbereidend beroepsonderwijs in vijftien jaar terug van 835 naar 82. Daarentegen groeide het aantal scholen met alle typen van voortgezet onderwijs, de zogeheten brede scholengemeenschappen. Het meest duidelijk wordt dat door te kijken naar het percentage leerlingen dat op een brede scholengemeenschap zit: in 1990 was dat 5 procent en in 1995 meer dan 50 procent. Bij deze schaalvergroting moet een duidelijke kanttekening worden gemaakt. Wat organisatie betreft zijn scholen de laatste jaren in omvang toegenomen, maar voor leerlingen is daar vaak weinig van te merken, want veel scholen bestaan uit verschillende locaties. Vandaar dat voor een effect op prestaties van leerlingen beter naar de locatiegrootte dan naar de schoolgrootte kan worden gekeken. De verwachting is dat de locatiegrootte negatief samenhangt met de prestaties van leerlingen. Een negatieve samenhang zou het gevolg kunnen zijn van de grootschaligheid op veel scholen, waarbij leerlingen mogelijk een ‘nummer’ worden en verloren gaan in de massa (Stoel, 1980; Hoffer & Coleman, 1990). In zijn column in de Volkskrant geeft Schuyt (1996) het als volgt weer: “Grote scholengemeenschappen verliezen het zicht op de leerlingen en de computer-gestuurde aanwezigheidslijsten geven wel een snelle uitdraai van spijbelende leerlingen, maar daarmee heb je ze nog niet terug in de klas. Persoonlijke binding helpt. (...) Wat in elk geval als kenmerk van grootschaligheid opvalt, is de geringere betrokkenheid van zeer velen, de sfeer van anonimiteit en onpersoonlijkheid, van aangemoedigde ontmoediging.”

Wat de leerlingencompositie betreft is met name een effect van het *percentage allochtone leerlingen* te verwachten. Scholen verschillen sterk in dit opzicht. Op de meeste scholen zitten niet of nauwelijks allochtone leerlingen, maar er zijn ook scholen die bijna volledig ‘zwart’ zijn.ⁱⁱ In hoofdstuk 2 heb ik laten zien dat deze segregatie sterk samenhangt met andere kenmerken van de leerlingencompositie. Scholen met veel allochtone leerlingen zijn doorgaans scholen waar ouders gemiddeld genomen een laag opleidingsniveau en leerlingen een laag instroomniveau hebben. Wanneer een leerling bijvoorbeeld op een locatie zit met veel leerlingen uit een bepaald sociaal milieu dan kan dit de keuze van vrienden beïnvloeden, wat van invloed kan zijn op de prestaties van leerlingen. Ook zou het kunnen zijn dat docenten op ‘zwarte’ scholen hun lessen aanpassen en lagere streefniveaus hanteren, omdat veel leerlingen taalproblemen hebben. Een dergelijke aanpassing kan een effect hebben op de prestaties van alle leerlingen.

De invloed van de *denominatie*, de levensbeschouwelijke richting, van de school is in hoofdstuk 1 reeds ter sprake gekomen. Sinds de jaren tachtig wordt er veelvuldig onderzoek gedaan naar de verschillen tussen openbaar en bijzonder onderwijs (Bryk, Lee & Holland, 1993; Coleman & Hoffer, 1987; Coleman, Hoffer & Kilgore, 1982). Deze onderzoekers concluderen dat katholieke scholen in de Verenigde Staten effectiever zijn dan andere scholen in de particuliere sector of scholen in de publieke sector. Dit succes van katholieke scholen kan grotendeels worden verklaard door verschillen in docent- en schoolkenmerken.

Met een aandeel van ongeveer 80 procent nemen bijzondere scholen in het Nederlandse voortgezet onderwijs een dominante positie in. In het bijzonder onderwijs zijn er twee grote religieuze denominaties, rooms-katholiek en protestant, die elk ongeveer een derde van het bestel vormen. Wanneer er rekening wordt gehouden met de effecten van kenmerken van leerlingen en gezinnen, is het de vraag of er prestatieverschillen tussen de denominaties zijn. De overheid bekostigt openbare en bijzondere scholen immers op gelijke wijze en geeft ook sturing aan beide sectoren. Het is dan ook minder gemakkelijk om een algemene uitspraak te doen over het succes van een bepaalde richting in het voortgezet onderwijs (Dijkstra, 1994; 1997).

Schoolkenmerken (cirkel B)

In de bespreking van de invloed van schoolkenmerken komen negen kenmerken aan bod die in de literatuur aan effectieve scholen worden toegeschreven, te weten nadruk op basisvaardigheden, professioneel en onderwijskundig leiderschap, eenheid in visie en doelen, ordelijk schoolklimaat, bewaking van leervorderingen, samenwerking tussen school en thuis, aard brugperiode, mentoraatsysteem, docent-leerling ratio en toelatingscriteria. In tegenstelling tot de contextuele ken-

merken kunnen scholen op deze kenmerken beleid voeren. Dergelijke kenmerken fungeren dan ook mogelijk als inspiratiebron voor scholen die hun kwaliteit willen vergroten (Van Petegem, 1997).

Tot de vijf factoren die volgens Edmonds (1979) de prestaties van leerlingen beïnvloeden behoort de *nadruk op basisvaardigheden*. Voor het Nederlandse basisonderwijs kunnen taal, rekenen en schrijven tot de basisvaardigheden worden gerekend. Voor het voortgezet onderwijs gaat het dan met name om Nederlands, Engels en wiskunde. De betere prestaties van leerlingen op katholieke scholen verklaren Bryk, Lee en Holland (1993) door de nadruk op basisvaardigheden. Katholieke scholen in de Verenigde Staten, zo stellen zij, zijn meestal vrij klein. Door het geringe leerlingenaantal hebben deze scholen minder mogelijkheden om allerlei keuzevakken en richtingen in te voeren. Vrijwel alle leerlingen volgen een academisch curriculum waarin de basisvaardigheden centraal staan, wat gemiddeld genomen een positief effect heeft op hun prestaties. Daarentegen zijn publieke scholen vaak groot, waardoor ze tal van extra vakken en aparte richtingen kunnen aanbieden. Gekscherend vergelijken Powell, Farrar en Cohen (1985, p. 309) publieke scholen met winkelcentra waar leerlingen kunnen bepalen wat ze willen kopen. “The mall's central qualities - variety of offerings, choice among them, and neutrality about their value - have succeeded in holding most teenagers on terms they and their teachers can live with. The will to learn is perceived, in a deceptively sensible formulation, simply as the responsibility of students and their families.” In tegenstelling tot de Verenigde Staten beschikt Nederland over een standaardcurriculum voor het voortgezet onderwijs met een grotendeels voorgeschreven lessentabel en een landelijk eindexamen. In het Nederlandse voortgezet onderwijs zijn er dus veel minder verschillen te verwachten in de mate waarin tijd wordt besteed aan de basisvaardigheden. Scheerens en Bosker (1997) stellen dan ook dat dit effectiviteitskenmerk typisch Amerikaans is.

Het professioneel en onderwijskundig leiderschap behoort ook tot de kenmerken die Edmonds in 1979 naar voren bracht. Onder een professioneel en onderwijskundig leiderschap wordt verstaan een directie die de structuur van en de relaties binnen de school bewaakt en die veel tijd en energie steekt in activiteiten om de school te verbeteren (Sammons, Hillman & Mortimore, 1995). Voor de structuur van de school is het belangrijk dat verschillende elementen van de school op elkaar aansluiten. De trefwoorden die Creemers (1991; 1994) hiervoor hanteert zijn consistentie, cohesie, constantie en controle. *Consistentie* houdt in dat de diverse componenten van het onderwijs op elkaar aansluiten, zoals de groeperingvorm, het onderwijsleermateriaal en het docentgedrag. Wanneer een leerling bij verschillende docenten en gedurende de gehele schooltijd op een zelfde manier leskrijgt dan is er sprake van *cohesie* en *constantie*. Door de leerresultaten

en het docentgedrag te evalueren kan de schoolleiding de effectieve instructie *controleren*. Deze trefwoorden sluiten aan bij een begrip dat Rutter e.a. (1979) hanteren. Verschillen tussen scholen hangen samen met *configuraties* van schoolkenmerken en niet zo zeer met specifieke schoolkenmerken. De schoolleiding moet dus nagaan of er succesvolle configuraties tot stand komen. Wat de omgang tussen docenten en directie betreft stellen Rutter e.a. dat het voor docenten belangrijk is om te merken dat ze deel uit maken van het beslissingsproces en dat ze worden betrokken bij beslissingen die voor hen van belang zijn. Het is volgens hen belangrijk dat er met docenten rekening wordt gehouden, maar wanneer nodig moeten de beslissingen hogerop worden genomen. In scholen die goed functioneren, zo vervolgen Rutter e.a., worden de belangrijke beslissingen genomen door de directie en niet door de docenten. Op deze manier schermt de directie het personeel af en vermijdt de leiding dat docenten te veel vergaderen. Wilson en Corcoran (1988) brengen naar voren dat een professionele leiding dient te beschikken over voldoende kennis van onderwijsonderzoek en van de praktijk in de school. Die kennis moet de directie inzetten om de school te verbeteren.

In tegenstelling tot het Angelsaksische onderzoek heeft het Nederlandse onderzoek naar een effect van het onderwijskundig leiderschap in het voortgezet onderwijs weinig empirische bevestiging opgeleverd (Scheerens & Bosker, 1997). Knuver (1993) meent dat de sterke rechtspositie van docenten een reden zou kunnen zijn voor het ontbreken van een effect van het leiderschap. De Nederlandse schoolleider heeft minder invloed op het aannemen of ontslaan van personeel dan hun collega's in andere landen, zoals een rector het treffend verwoordt (Drayer, 1997, p. 23). “Echt personeelsbeleid voeren kun je niet. (...) We zullen komend jaar tien leerkrachten moeten overplaatsen. En dat zijn de jonge mensen, want die zijn het laatst aangekomen. (...) Juist docenten die vitaal zijn, die nog vol *power* zitten, raak ik kwijt. Een man van zesenvijftig die graag weg wil, die is opgebrand, kan niet wegkomen. Absurd. De doorstroming is nihil. Eigen keuze heb ik niet. Als ik al een vacature heb, moet ik eerst wachtgelders aannemen. Zo kun je nooit onderwijskundige vernieuwingen doorvoeren.” Het is dus de vraag of directies van Nederlandse scholen voor voortgezet onderwijs op dit aspect de regie stevig in handen kunnen hebben.

Op scholen waar onder personeel een *eenheid in visie en doelen* leeft, zo komt uit onderzoek van Bryk, Lee & Holland (1993) naar voren, zijn de prestaties van leerlingen beter. Ook Chubb en Moe (1990, p. 91) benadrukken het belang van een gezamenlijke visie en doelen. Op effectieve scholen is volgens hen de collegialiteit onder docenten hoger. “(T)eachers in the best schools are especially likely to agree that their fellow teachers 'can be counted on anytime, anywhere'; that they all 'share beliefs in the central school mission'; and among other things, that the school is 'like

a big family'." In Nederland vermoeden verschillende wetenschappers, zie bijvoorbeeld Meijnen e.a. (1991), het belang van dit kenmerk voor prestaties van leerlingen, maar tot empirisch onderzoek heeft dit nog niet geleid.ⁱⁱⁱ Overigens leiden niet alle onderzoeken naar eenheid in visie en doelen tot empirische ondersteuning. Phillips (1997, p. 635) vindt in haar onderzoek geen effect van wat zij communitarisme noemt. Zij noemt het onderzoek naar een kenmerk als eenheid in visie en doelen tijdsgevoelig. "In this postindustrial era of geographic mobility and rapidly changing family structures, many people believe that traditionally communal domains, such as neighborhoods and families, have become less stable and supportive than they once were. Communitarian theories suggest that teachers' disaffection and students' school failure arise from a lack of strong affective ties to both places and people; a void that communally organized schools may fill."

De geschiktheid van de omgeving voor het leren is een schoolkenmerk waaraan onderzoekers al sinds geruime tijd aandacht besteden. In onderzoek naar zowel schooleffectiviteit (Edmonds, 1979; Rutter e.a., 1979; Wilson & Corcoran, 1988) als verschillen tussen openbaar en bijzonder onderwijs (Bryk, Lee & Holland, 1993; Coleman & Hoffer, 1987; Coleman, Hoffer & Kilgore, 1982) komt een *ordelijk schoolklimaat* als belangrijk naar voren. Op de meeste effectieve scholen heerst volgens Sammons, Hillman en Mortimore (1995) voldoende orde om het lesgeven en leren ongestoord door te laten gaan. Een van de middelen die de directie kan aanwenden om het lesgeven zo veel mogelijk door te laten gaan is het invoeren van een invalregeling bij afwezigheid van een docent. Op sommige scholen blijken docenten op hun rooster een of twee uur te hebben staan waarop ze verplicht kunnen worden om in te vallen voor een collega (Veenstra, 1997). De ordelijkheid van de leeromgeving is mogelijk meer van belang voor leerlingen in lagere dan voor leerlingen in hogere vormen van voortgezet onderwijs. Wanneer de regels en afspraken op school niet al te straf zijn, gaan die leerlingen eerder over tot bijvoorbeeld spijbelen (vergelijk Rutter e.a., 1979).

De voortgang van leerlingen in prestaties dient volgens Edmonds (1979) niet alleen op klasniveau maar ook op schoolniveau te worden bewaakt. Om de *leervorderingen* te *bewaken* kunnen scholen een systeem opzetten waarin ze de voortgang registreren. In een dergelijk systeem houden scholen van iedere leerling gegevens bij over bijvoorbeeld de thuissituatie, de inzet, het gedrag, schoolverzuim, de proefwerkcijfers, de rapporten, de vakkenpakketkeuze en de wensen van vervolgopleiding. Van der Werf, Van der Tuin en Driessen (1995) wijzen erop dat een minderheid van de scholen zo uitgebreid over leerlingen rapporteren. In Scheerens en Bosker (1997) zijn geen aanwijzingen te vinden dat er voor dit kenmerk empirische steun is in onderzoek naar het Nederlandse voortgezet onderwijs.

De empirische onderbouwing van de invloed *samenwerking tussen de school en de ouders* van leerlingen op leerprestaties is volgens Scheerens (1992) vrij stevig. Die steun wordt echter verkregen door onderzoek in het basisonderwijs, waar het geregeld voorkomt dat ouders helpen bij bijvoorbeeld het lezen of buitenschoolse activiteiten. Sammons, Hillman en Mortimore (1995) noemen in hun overzichtsartikel dan ook alleen maar studies voor het basisonderwijs. In het voortgezet onderwijs blijken deze contacten tamelijk formeel (Veenstra, 1997). Een paar keer per jaar organiseert de school spreekuren voor ouders. Verder zijn er jaarlijks enkele thema-avonden om ouders voor te lichten over de keuze van een vakkenpakketten of vervolgopleidingen. Voor veel ouders zijn dat de weinige momenten waarop ze contact hebben met de school. Voor het voortgezet onderwijs is er voor het effect van dit schoolkenmerk weinig empirisch houvast (vergelijk Ho & Willms, 1996).

Wanneer er in het voortgezet onderwijs wordt gedifferentieerd, gaat het doorgaans om niveaudifferentiatie tussen klassen. Tussen scholen kan de mate van differentiatie, *de aard van de brugperiode*, variëren. Op sommige scholen worden leerlingen op basis van hun prestatieniveau in zo homogeen mogelijke prestatiegroepen geplaatst. Op andere scholen worden klassen juist heterogeen samengesteld. Oakes (1985) heeft een beschrijvend onderzoek gedaan naar de invloed van heterogeen en homogeen groeperen. In haar conclusie houdt zij een pleidooi voor heterogeen groeperen met als argument dat geen enkele groep van leerlingen volgens haar profiteert van het krijgen van les in een homogene groep. Daarentegen vinden Dar en Resh (1986) in hun onderzoek naar het voortgezet onderwijs in Israël wel een effect van de groeperingvorm. Ook zij zijn voorstander van heterogeen groeperen, want bij die vorm van groeperen is de achteruitgang in prestaties van goede leerlingen kleiner dan de vooruitgang van zwakke leerlingen.^{iv} Gamoran (1992, p. 826) gaat in op verschillende aspecten van homogeen en heterogeen groeperen. Zijn conclusie is dat naarmate de klassensamenstelling van een school voor langere tijd vast ligt, de prestatieverschillen tussen klassen groter zijn en de leerprestaties voor alle klassen te zamen gemiddeld genomen lager zijn. “(T)he results for track immobility suggest that inflexible tracking systems have greater inequality along with lower average achievement, presumably reflecting especially poor conditions in the nonacademic tracks of such schools.” In Nederland heeft De Vries (1992) onderzoek gedaan naar de klassensamenstelling van scholen voor voortgezet onderwijs. Haar conclusie sluit aan bij die van Oakes. Op termijn blijkt er geen invloed van de klassensamenstelling op de onderwijspositie van leerlingen. Toch prefereert zij de heterogene groeperingvorm.

Scholen blijken behoorlijk te verschillen in de mate waarin ze aandacht besteden aan *leerlingbegeleiding* in het algemeen en het *mentoraatsysteem* in het bijzon-

der (Veenstra, 1998). Op minder dan 10 procent van de scholen blijkt het mentoraatsysteem uitgebreid opgezet, wat inhoudt dat zowel het aantal mentorlessen als de frequentie van overleg tussen mentoren hoog is. Op deze scholen hebben leerlingen in de eerste leerjaren in ieder geval één en vaak twee of drie mentoruren per week. Niet alleen voor de leerlingen maar ook voor de docenten is het mentoraat in dat systeem intensief, want op deze scholen overleggen mentoren doorgaans wekelijks. Daartegenover staat een grote groep van scholen, bijna viertiende van de scholen voor voortgezet onderwijs, waar het mentoraatsysteem beperkt is qua omvang. Op deze scholen hebben leerlingen meestal alleen in de brugklas een mentoruur en overleggen mentoren eens in de drie maanden. De overige scholen nemen een tussenpositie in. Naar de invloed van het mentoraatsysteem op prestaties van leerlingen is weinig onderzoek gedaan. Er zijn slechts drie studies bekend. Van Marwijk Kooy-von Baumhauer (1984) stelt dat scholen die een actief beleid voor begeleiding van leerlingen voeren goede examenresultaten behalen. Tesser (1986) vindt op het vwo een positieve correlatie tussen de hoeveelheid leerlingbegeleiding en het bereikte onderwijsniveau na vijf jaar in het voortgezet onderwijs. Van der Tuin (1993) laat voor scholengemeenschappen zien dat er een positieve bijdrage uitgaat van leerlingbegeleiding op de prestaties van leerlingen. Die positieve bijdrage zou kunnen komen doordat scholen voortijdig inspelen op mogelijke problemen van leerlingen. Tijdens de mentorlessen kan bijvoorbeeld aandacht worden besteed aan studie- en sociale vaardigheden.

Een grote school hoeft niet kil en onpersoonlijk te zijn. Sizer (1992) doet hiertoe de aanbeveling om de *docent-leerling ratio* te verkleinen. Hij stelt dat een verlaging van deze ratio mogelijk is zonder dat dit extra kosten met zich mee brengt. Zijn voorstel is om docenten te groeperen in clusters, en om elke docent meer dan een vak te laten geven. Vertaald naar de Nederlandse situatie, zouden geschiedenis, aardrijkskunde, maatschappijleer en economie te zamen een vak kunnen vormen, bijvoorbeeld onder de noemer 'mens en maatschappij'. Natuurkunde, scheikunde en biologie kunnen eveneens tot een geheel worden gesmeed. Ook bij de talen zou meer kunnen worden gecombineerd. Vaardigheden die bij alle talen van belang zijn, zoals spreekvaardigheid, zouden op een standaardmanier kunnen worden geoefend. Het ideaal van Sizer is dat docenten aan niet meer dan tachtig leerlingen lesgeven, zodat een persoonlijke band tussen docenten en leerlingen ook in het voortgezet onderwijs mogelijk wordt. Scholen waar docenten aan een dergelijk aantal leerlingen lesgeven, zijn er in Nederland nauwelijks. Het enige reguliere schooltype waar een dergelijke omvang wordt benaderd is het ivbo. Daar is de grootte van een klas maximaal vijftien leerlingen. In het ivbo geven docenten vaak verschillende vakken aan dezelfde groep leerlingen. De docent-leerlingratio is er dan ook laag. Wellicht dat een lage ratio

ook van belang is voor leerlingen van andere schooltype en dan met name vbo-leerlingen.

Door de selectiemechanismen die katholieke scholen in de Verenigde Staten toepassen behalen ze, zo beargumenteert Witte (1992), betere prestaties dan andere scholen. Zij hanteren volgens hem strengere *toelatingscriteria* dan publieke scholen en weren daardoor probleemrijke leerlingen. Katholieke scholen eisen van leerlingen niet alleen dat ze het niveau van de school aankunnen, maar ook dat de leerlingen binnen het schoolklimaat passen. In Nederland hanteren scholen als toelatingseis het advies van de basisschool. Sommige scholen willen daarnaast ook het prestatieniveau van een leerling weten. Mogelijk dat scholen op die manier strenger selecteren en daardoor een makkelijkere leerlingenpopulatie krijgen. Het prestatieniveau van een leerling en het advies van de basisschool hangen weliswaar sterk samen, maar deze samenhang is niet perfect. Zo krijgen allochtone leerlingen en leerlingen uit hogere sociale milieus hogere adviezen dan hun score op de Cito-toets aangeeft (Dronkers, 1997a).

Docentkenmerken (cirkel A)

Tot slot bespreek ik twee kenmerken van effectieve docenten, namelijk gelegenheid tot leren en effectieve leertijd en gestructureerd onderwijs.^v

De *gelegenheid tot leren* is een term die Carroll (1963) heeft ingevoerd. In een klassiek artikel werkt hij een model van de mate van leervorderingen uit. Dat model bestaat uit vijf elementen die zijn verdeeld over twee dimensies. De eerste dimensie is de daadwerkelijk bestede leertijd, die volgens Carroll gelijk is aan het minimum van de hoeveelheid tijd die nodig is om een taak onder ideale omstandigheden te beheersen (het leertempo van een leerling), de toegestane leertijd (de gelegenheid tot leren) of de tijd die leerlingen bereid zijn actief aan leren te besteden (het doorzettingsvermogen van een leerling). De benodigde leertijd vormt de tweede dimensie en is, zo vervolgt hij, gelijk aan het leertempo van leerlingen onder controle van de kwaliteit van de instructie en hun begripsvermogen. De mate van vooruitgang in prestaties is gelijk aan de quotiënt van de daadwerkelijk bestede en de benodigde leertijd (zie voor de formule: Carroll, 1962). Wanneer deze quotiënt groter is dan één, dan presteren leerlingen boven hun niveau en wanneer de quotiënt kleiner is dan één, dan presteren leerlingen onder hun niveau.

Van de genoemde kenmerken van het model zijn de gelegenheid tot leren, de kwaliteit van de instructie en het doorzettingsvermogen van de leerling mogelijk te beïnvloeden (Carroll, 1963). Volgens Carroll (1989) is het vooral belangrijk om er achter te komen hoe de gelegenheid tot leren kan worden vergroot. Om de toegestane leertijd te bepalen is het begrip *effectieve leertijd* van belang. De effectieve leertijd is de tijd die per lesuur resteert als onderbrekingen door orde- of

concentratieproblemen niet meetellen (vergelijk Leseman, Sijsling & De Vries, 1992). De effectieve leertijd is afhankelijk van (1) de ingeroosterde lestijd, (2) de orde en rust in de klas, (3) de prestatiemotivatie van leerlingen, (4) de taakgerichtheid en het planningsgedrag van docenten en (5) de extra leertijd door opgave en controle van huiswerk.

Om de effectieve leertijd te vergroten is het volgens Creemers (1991) belangrijk dat docenten leerlingen bij de les houden door het geven van *gestructureerd onderwijs*. Leerlingen leren volgens Veenman (1992) meer, geordender en efficiënter als de docent in het onderwijs die zaken aan de orde stelt die de leerling nog niet begrijpt of beheerst en die liggen in, wat de leerpsycholoog Vygotsky noemt, hun zone van naaste ontwikkeling. Het onderwijs moet ontwikkelend en niet volgend zijn. Hierbij is het volgens Brophy en Good (1986, p. 361) van belang dat docenten de activiteiten van leerlingen sturen. "Students achieve more in classes where they spend most of their time being taught or supervised by their teachers rather than working on their own." Leerlingen krijgen dan minder mogelijkheden om te verslapen tijdens de les. Gestructureerd onderwijs, dat ook wel bekend staat als directe instructie of sturend onderwijs, bestaat volgens Doyle (1985) uit een achttal aspecten.^{vi} Allereerst is het van belang dat een docent de leerdoelen van een les duidelijk formuleert en precies bedenkt wat hij de leerlingen wil leren. Vervolgens is het nodig dat hij de leerstof opsplitst in voor de leerling overzienbare eenheden. De docent moet met kleine stappen door de stof gaan, waarbij hij deze stappen in een weloverwogen volgorde aanbiedt. In de derde plaats moet hij bij de aanvang van de les de leerlingen duidelijk laten weten wat de doelstellingen zijn. Daarna moet de docent eerder geleerde stof herhalen, opdat hij de benodigde voorkennis van leerlingen ophaalt. Veenman (1992) stelt dat een docent die bij het begin van zijn les duidelijk aangeeft wat de doelen zijn en die een samenvatting geeft van de voorafgaande leerstof, de leerlingen helpt om betekenis te geven aan de nieuwe leerstof. In de vijfde plaats moet de docent duidelijk uitleg geven. Hij moet heldere taal gebruiken, concrete voorbeelden geven, hoofdzaken samenvatten en geregeld vragen stellen. Als zesde punt geldt dat leerlingen voldoende tijd moeten krijgen om de stof te oefenen. Deze oefening dient eerst onder begeleiding en daarna zelfstandig te gebeuren. Het is hierbij van belang dat leerlingen weten dat docenten hun werk controleren en dat ze deze controle zo spoedig mogelijk uitvoeren. Op die manier merken leerlingen gelijk de invloed van het nakijken door de docent. Bij deze terugkoppeling moeten ze ook aangeven waarom iets goed of fout is. Een volgend aspect is de frequente evaluatie van de voortgang van leerlingen door de docent. De kennis, inzicht en vaardigheden van leerlingen moeten geregeld worden getoetst, omdat een docent er zo achter kan komen of de leerlingen de stof beheersen. Wanneer leerlingen de leerdoelen niet bereiken, moet hij daar daadwerkelijk iets aan

doen door bijvoorbeeld zijn lessen bij te sturen. Het laatste kenmerk van gestructureerd onderwijs is de periodieke herhaling van de leerstof met als doel dat de leerlingen de leerstof automatisch gaan beheersen. De kern van deze instructiemethode is volgens Rosenshine en Stevens (1986) de opeenvolging van de uitleg van de docent, het oefenen door de leerlingen en de corrigerende terugkoppeling door de docent. Op deze manier geven docenten richting aan het leerproces van leerlingen.

Besluit

De empirische steun voor de invloed van schoolkenmerken op prestaties van leerlingen is doorgaans zwak of afwezig. Het is dan ook twijfelachtig of de genoemde kenmerken een effect hebben. Verscheidene kenmerken zijn eerder van kracht in het basis- dan in het voortgezet onderwijs. Ook geldt dat enkele in het buitenland gevonden verbanden, zoals de nadruk op basisvaardigheden en het professioneel en onderwijskundig leiderschap, in Nederland nauwelijks of niet worden teruggevonden. Verder geldt dat bij de empirische steun voor verschillende kenmerk kanttekeningen zijn te plaatsen. Wanneer onderzoekers een adequaat design gebruiken, een multiniveau analyse naar verschillen in netto prestaties, dan worden er minder sterke effecten gevonden dan in onderzoek op één niveau of in onderzoek naar verschillen in bruto prestaties (Scheerens & Bosker, 1997).

Voor de invloed van klassenkenmerken op prestaties van leerlingen is de empirische steun vrij stevig (Creemers, 1991). Effecten van gelegenheid tot leren en gestructureerd onderwijs worden geregeld aangetoond. De meeste van deze effecten zijn echter nog niet in onderzoek naar het voortgezet onderwijs in Nederland gevonden.

De verwachting is dat in de onderbouw van het voortgezet onderwijs in Nederland geldt dat de netto prestaties van scholen hoger zijn, wanneer: (1) de locatiegrootte kleiner is; (2) het percentage allochtone leerlingen lager is; (3) de eenheid in visie en doelen groter is; (4) het schoolklimaat ordelijker is; (5) het mentoraatsysteem intensiever is; (6) de docent-leerling ratio kleiner is; (7) de toelatingscriteria strenger zijn; (8) de gelegenheid tot leren groter is; (9) het onderwijs meer gestructureerd is. Verder wordt getoetst of er verschillen zijn in netto prestaties tussen scholen (10) van verschillende denominaties; (11) met verschillende vormen van brugperiode. In Nederland is het onduidelijk in welke richting deze effecten zullen optreden.

Deze effecten hoeven niet voor alle schooltypen hetzelfde te zijn. Voor het effect van bijvoorbeeld de ordelijkheid van het schoolklimaat of de docent-leerling ratio zou kunnen gelden dat het effect vooral voor het beroepsonderwijs opgang

doet. Over dergelijke verschillen in effecten tussen hogere en lagere vormen van voortgezet onderwijs is uit de literatuur overigens weinig bekend.

Van sommige kenmerken worden in het voortgezet onderwijs geen effecten verwacht, te weten de nadruk op basisvaardigheden, het professioneel en onderwijskundig leiderschap, de bewaking van leervorderingen en de samenwerking tussen school en thuis.^{vii}

4.3 Meten van docent- en schoolkenmerken

In deze paragraaf ga ik allereerst in op de respons, de operationalisatie en ontbrekende waarden van kenmerken op klas- en schoolniveau. Vervolgens geef ik de beschrijvende statistieken.

Respons

In dit onderzoek maak ik gebruik van vragenlijsten voor de directie en voor docenten. Van de directieleden zijn voor het schooljaar 1993-'94 de gegevens beschikbaar van 142 van de 149 scholen (95 procent) die ik mijn analyses heb opgenomen. Voor 1995-'96 zijn de gegevens van 111 van de 149 scholen (74 procent) beschikbaar. Van 471 docenten van 124 scholen zijn gegevens beschikbaar. Aangezien er in dit onderzoek 453 klassen zijn en docenten voor Nederlands en wiskunde zijn ondervraagd, hadden er 906 docenten een vragenlijst kunnen inleveren. De respons bedroeg dus 52 procent. In 1995-'96 is informatie verkregen van 522 docenten van 135 scholen. Op docentniveau is de respons 58 procent.

Operationalisatie

De contextuele schoolkenmerken zijn afkomstig uit de directievragenlijst. Bij de *denominatie* onderscheid ik zeven typen: (1) openbaar; (2) algemeen bijzonder; (3) interconfessioneel (inclusief samenwerkingsscholen); (4) protestants-christelijk; (5) rooms-katholiek; (6) reformatorisch; (7) gereformeerd-vrijgemaakt. Verder neem ik het *percentage allochtone leerlingen*^{viii} en de natuurlijke logaritme van de *locatiegrootte* in de analyses mee. In onderwijsonderzoek is het tamelijk gebruikelijk om de \log_{10} van de locatiegrootte te nemen, omdat het verband tussen de locatiegrootte en de netto prestaties zo het beste wordt weergegeven.

Tenzij anders vermeld zijn de schoolkenmerken ook afkomstig uit de directievragenlijst uit het schooljaar 1993-'94. Om de *eenheid in visie en doelen* te meten gebruik ik gegevens uit de docentenvragenlijst uit 1995-'96. Aan docenten zijn twaalf vragen gesteld over de eenheid binnen de school, zie Bijlage 6. Van deze vragen heb ik de standaarddeviatie gebruikt om de spreiding in de antwoorden van do-

centen te meten. Aan de docenten is in 1995-'96 gevraagd hoe lang ze al op hun school werken. Met die vraag kan de teamstabiliteit worden bepaald. De teamstabiliteit zegt mogelijk ook iets over de eenheid binnen een school. Voor de operationalisatie van een *ordelijke leeromgeving* gebruik ik gegevens over afspraken tussen de directie en docenten, zie Bijlage 6. De vragen vormen te zamen een schaal met een Cronbachs alfa van 0,74. Ook gebruik ik gegevens over de actie die de directie onderneemt als een les onverwacht uitvalt. Hierbij onderscheid ik drie categorieën: (1) huiswerk maken; (2) invalregeling; (3) lesuur vrij of een andere maatregel. Bij *de aard van de brugperiode* onderscheid ik drie vormen, te weten homogeen, dakpansgewijs en heterogeen. Homogeen is een klas, wanneer er alleen leerlingen van één schooltype bijeenzitten. Dakpansgewijs is een klas, wanneer er leerlingen van twee schooltypen bijeenzitten, bijvoorbeeld een vbo/mavo-klas, en heterogeen is een klas bij leerlingen van drie of meer schooltypen. Vervolgens is per school berekend hoeveel leerlingen er in elk van de groeperingvormen zitten. Sommige scholen hebben alle leerlingen op een zelfde wijze gegroepeerd. Andere scholen hanteren verschillende vormen naast elkaar. Deze scholen krijgen dan een score op meer dan één groeperingvorm. Bij het *mentoraatsysteem* (gemeten in 1995-'96) onderscheid ik vier categorieën: (1) smal-verticaal; (2) smal-horizontaal; (3) breed-verticaal; (4) breed-horizontaal. Scholen met een uitgebreid mentoraatsysteem noem ik breed-horizontaal. Dit zijn scholen waar in de eerste vier leerjaren gemiddeld meer dan één mentoruur per jaar staat ingeroosterd en waar mentoren vrijwel wekelijks vergaderen. Scholen met een beperkt mentoraatsysteem noem ik smal-verticaal (Veenstra, 1998). Op deze scholen hebben leerlingen alleen in de brugklas of helemaal geen mentoruur. Mentoren vergaderen op deze scholen uit eens per kwartaal of minder vaak. In de andere twee categorieën scoren scholen hoog op het ene en laag op het andere kenmerk. Smal-horizontale scholen zijn scholen waar een mentoruur niet al te vaak voorkomt, hoewel een mentorenoverleg maandelijks op het rooster staat. Het omgekeerde geldt voor breed-verticale scholen. Op deze scholen komt in de meeste leerjaren een mentoruur voor, maar mentoren overleggen gemiddeld niet veel vaker dan eens in de drie maanden. De *docent-leerling ratio* heb ik berekend door met gegevens uit 1995-'96 voor docenten na te gaan aan hoeveel leerlingen zij lesgeven en dat totaal te delen door het aantal uren dat ze wekelijks lesgeven. Als een docent met 24 lesuren aan twaalf klassen van 25 leerlingen lesgeeft, dan is de ratio: $12 * 25 / 24 = 12,5$. Vervolgens heb ik per school het gemiddelde van deze ratio's berekend. Behalve het advies van de basisschool kunnen scholen het prestatieniveau van leerlingen als *extra toelatingscriterium* hanteren. Deze variabele is een dichotome variabele. Scholen hanteren dit extra criterium wel of niet.

Voor de operationalisatie van de docentkenmerken gebruik ik gegevens uit de vragenlijsten voor de docenten Nederlands en wiskunde die leerlingen in de brugklas hadden (afgenomen in schooljaar 1993-'94). Voor de operationalisatie van de *effectieve leertijd* gebruik ik een vraag over de tijdsindeling van docenten. De *effectieve leertijd* is het percentage van de totale lestijd dat aan instructie (uitleg van nieuwe leerstof en onder begeleiding oefenen van nieuwe stof), verwerking (individueel of in groepjes laten werken aan opdrachten of huiswerk) en controle (bespreken of overhoren van huiswerk of in de klas gemaakte opdrachten), maar niet aan ordehandhaving wordt besteed. Om na te gaan of docenten *gestructureerd onderwijst* geven heb ik geprobeerd verschillende vragen te combineren. Een docent geeft gestructureerd les als de les bestaat uit de opeenvolging van uitleg door de docent, oefenen door leerlingen en corrigerende terugkoppeling door de docent. In de docentenvragenlijsten zijn er verschillende schalen voor de verschillende onderdelen van lesgeven. In Bijlage 6 worden deze schalen vermeld. De schalen blijken betrouwbaar. De betrouwbaarheid is voor alle schalen hoger dan 0,80. De schalen blijken onderling ook vrij sterk te correleren. De correlaties variëren van 0,40 tot 0,77. Uit een nadere inspectie van de data valt echter de nodige twijfel bij deze docentgegevens te plaatsen, zie Tabel 4.1.

Tabel 4.1: Enkele gegevens uit het bestand met docentkenmerken

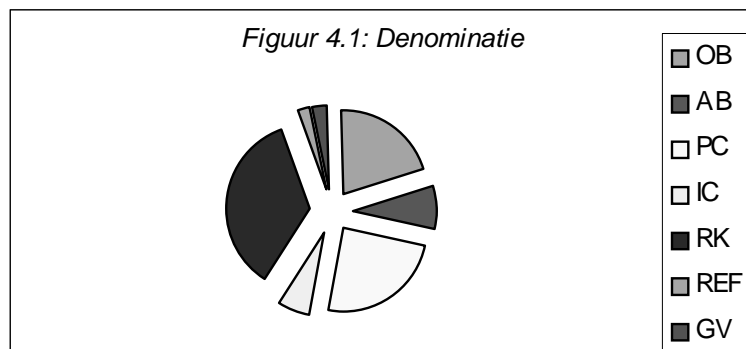
NUMMER	VR36A	VR36B	VR36C	VR36D	VR36E	VR36F
1	100	100	100	100	100	100
2	75	50	20	75	20	25
3	50	50	10	10	25	5
4	20	999	999	60	20	999
5	90	30	50	50	20	30
6	999	999	999	999	999	999
7	30	20	15	15	10	10
8	999	999	999	999	999	999
9	10	10	30	20	30	20
10	20	20	10	10	5	10

In Tabel 4.1 zijn voor tien docenten de antwoorden op de schaal 'Duidelijk en gestructureerd uitleggen' weergegeven. Het blijkt dat docenten de vragen op verschillende manieren benaderen. Sommige docenten hebben geen zin om deze vragen in te vullen en hebben talrijke ontbrekende waarden. In de tabel worden ontbrekende waarden met het getal 999 aangeduid. Docenten 6 en 8 hebben geen enkele vraag beantwoord. De andere docenten hebben één of meer vragen beant-

woord. Docent 1 legt naar eigen zeggen erg duidelijk en gestructureerd uit. Op alle vragen is het antwoord de maximale score van 100. Daartegenover staat docent 10 wiens antwoorden variëren van 5 tot 20. Tussen de scores van deze twee docenten in, bevinden zich de scores van bijvoorbeeld docent 2. Op sommige vragen scoort deze docent hoog en op andere vragen laag. Los van de ontbrekende waarden is er tot dusverre niet zo veel aan de hand met deze gegevens. Het probleem is echter dat voor sommige docenten de schaal van 0 tot 100 niet per vraag geldt, maar voor alle vragen gezamenlijk. De somscores van de zes vragen is voor docenten 4 en 7 gelijk aan 100. Voor docenten 9 en 10 is onduidelijk welke schaal ze hebben gehanteerd. Hebben ze per vraag een schaal van 0 tot 100 gehanteerd en scoren ze dus laag wat duidelijk en gestructureerd uitleggen betreft? Of hebben ze ook een schaal van 0 tot 100 voor alle vragen gezamenlijk gehanteerd en hebben ze een rekenfout gemaakt? Deze laatste mogelijkheid is reëel, omdat de omvang van de docentenvragenlijst zonder meer gevolgen moet hebben gehad voor de nauwkeurigheid in de antwoorden van docenten. De slotsom is dus dat de data betrouwbaar, maar niet valide zijn. Vandaar dat in het vervolg van dit onderzoek geen gebruik wordt gemaakt van docentgegevens. De vraag naar effecten van docentkenmerken kan daarom niet worden beantwoord.

Ontbrekende waarden

Ontbrekende waarden komen in de vragenlijsten voor directieleden voor. Deze gegevens heb ik steeds zo goed mogelijk proberen aan te vullen. Een contextueel kenmerk als de denominatie van de school heb ik *opgezocht*. Andere gegevens heb ik vervangen met zogeheten *regressie-imputatie* (Huisman, 1999). Om een ontbrekende waarde bij het percentage allochtone leerlingen in het schooljaar 1993-'94 te vervangen, heb ik een regressieanalyse van hetzelfde percentage voor het schooljaar 1995-'96 op het percentage van twee jaar eerder verricht. De ontbrekende waarden heb ik vervolgens vervangen door de waarde die door de regressievergelijking wordt voorspeld. Als deze methode niet kon worden gehanteerd, heb ik gekozen voor *gemiddelde-score-imputatie*.



Beschrijvende statistieken

Van de 149 scholen uit dit onderzoek is 21 procent openbaar (OB), 24 procent protestants-christelijk (PC) en 37 procent katholiek (RK). De overige scholen maken deel uit van een van de vier andere richtingen die ik onderscheid (AB: Algemeen bijzonder; IC: Interconfessioneel; REF: Reformatorisch; GV: Gerformeerd-vrijgemaakt). In Figuur 4.1 zijn de denominaties weergegeven met de richting van de wijzers van een klok.

Scholen variëren sterk wat locatiegrootte betreft. De gemiddelde grootte is 697 leerlingen, maar het bereik loopt van 124 tot 2.380 leerlingen. Het percentage allochtone leerlingen verschilt ook sterk van school tot school. Het gemiddelde is 7 procent, met als bereik 0 tot 87 procent (zie Tabel 4.2).

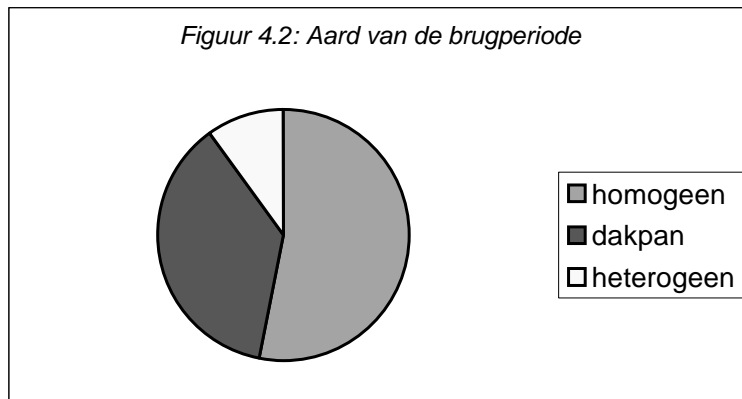
Tabel 4.2: Overzicht van schoolkenmerken

	GEM.	SD	MIN.	MAX.	N
CIRKEL C: CONTEXTUELE KENMERKEN					
Locatiegrootte	697	444	124	2.380	149
Percentage allochtone leerlingen	0,07	0,11	0,00	0,87	149
CIRKEL B: SCHOOLKENMERKEN					
Ordelijk schoolklimaat					
- afspraken over orderegels (4 items)	14,5	3,0	6,0	20,0	149
Eenheid in visie en doelen					
- standaarddeviatie (12 items)	9,8	0,9	5,7	12,4	149
Teamstabiliteit	15,0	3,6	7,2	25,3	149
Docent-leerling ratio	8,8	2,1	4,2	13,6	149

Directieleden scoren op de vier vragen over afspraken over orderegels gemiddeld 14,5, zie Tabel 4.2. Op scholen zijn er doorgaans wel afspraken vastgelegd over orderegels en docenten kunnen daarvan gemiddeld genomen alleen afwijken na overleg met de directie. Wat de maatregelen bij lesuitval betreft geven 83 van de 149 scholen aan te beschikken over een invalregeling, 21 scholen laten leerlingen tijdens het vrijgekomen uur huiswerk maken en 45 scholen laten dat lesuur vervallen.

De spreiding in de antwoorden van docenten over de eenheid binnen de school is 9,8. Op scholen waar deze spreiding gering is, zijn docenten het meer met elkaar eens en op scholen waar docenten het meer met elkaar oneens zijn is de spreiding groter. De Cronbachs alfa van de schaal is 0,74, zie Bijlage 6. Het aantal jaren dat docenten van een school gemiddeld genomen werkzaam zijn, varieert van 7,2 tot 25,3 jaar. Het gemiddelde bedraagt 15,0.

Wat de aard van de brugperiode betreft kunnen scholen verschillende scores tegelijkertijd hebben. Sommige scholen stellen alle klassen op een zelfde wijze samen. Het bereik op deze variabelen loopt dan ook van komt niet voor (0 procent) tot komt in alle klassen voor (100 procent). Gemiddeld zit 53 procent van de leerlingen in een homogene, 37 procent in een dakpansgewijze en 10 procent in een heterogene brugklas, zie Figuur 4.2.^{ix}



Het mentoraatsysteem is op 54 scholen smal-verticaal, op 46 scholen smal-horizontaal, op 34 scholen breed-verticaal en op 15 scholen breed-horizontaal. De docent-leerling ratio is 8,8. De scores op deze variabele lopen uiteen van 4,2 tot 13,6. Uitgaande van 28 lessen zijn er dus scholen waar docenten gemiddeld genomen aan ongeveer 120 leerlingen lesgeven. Dit minimum is nog hoger dan wat Sizer (1992) ideaal acht, namelijk 80 leerlingen per docent. Er zijn echter ook scholen waar docenten aan ruim drie keer zoveel leerlingen lesgeven. Bij 28 lessen geeft de gemiddelde docent les aan 250 leerlingen. Tot slot wordt het prestatieniveau op 78 procent van de scholen als toelatingscriterium gehanteerd.

4.4 Controles op klas- en schoolniveau

In deze paragraaf beantwoord ik de eerste drie deelvragen. Als eerste ga ik na in hoeverre het schooltype extra variantie verklaart in het spaarzame model met leerling- en gezinskenmerken uit het vorige hoofdstuk. Als tweede onderzoek ik of de effecten van leerling- en gezinskenmerken voor alle leerlingen gelijk zijn of dat er sprake is van differentiële effecten. Hierbij wordt gekeken naar random hellingen en intercept-helling covarianties. Als derde ga ik na of er regressie-effecten tussen klassen en scholen zijn.

4.4.1 Belang van het schooltype

Door de toevoeging van het door een leerling gevolgde schooltype van leerlingen in leerjaar drie wordt extra informatie over leerlingen in de analyses opgenomen. Om scholen eerlijk met elkaar te kunnen vergelijken, is het van belang om zoveel mogelijk met verschillen op individueel niveau rekening te houden.^x In hoofdstuk 3 heb ik talrijke leerling- en gezinskenmerken opgenomen, waaronder de eerdere prestaties van leerlingen. Die prestaties uit leerjaar één verklaren de prestaties uit leerjaar drie uiteraard goed. Naarmate brugklassers beter presteerden op de eerdere prestatietoetsen, scoren ze twee jaar later gemiddeld genomen hoger op Nederlands of wiskunde. Voor de wiskundeprestaties is de rekentoets de beste voorspeller. De verklaring van de wiskundeprestaties verbetert als de scores op de twee andere toetsen ook in de analyse worden opgenomen. Hoe meer er over het beginniveau van een leerling bekend is, hoe beter de verklaring. Door de opname van het schooltype dat leerlingen in het derde leerjaar volgen wordt de verklaring van hun scores aan het einde van dat leerjaar op Nederlands en wiskunde waarschijnlijk verbeterd.

Het schooltype is een ordinale variabele met zeven categorieën: (1) ivbo; (2) vbo; (3) vbo/mavo; (4) mavo; (5) havo; (6) havo/vwo; (7) vwo. Om het effect van schooltype weer te geven, is een model opgesteld met zes dummy's. In dit model vormen mavo-leerlingen de referentiegroep. De effecten van de overige schooltypen worden dus tegen die van mavo-leerlingen afgezet. Hierbij is de verwachting dat voor leerlingen op een schooltype hoger dan mavo geldt dat de regressiecoëfficiënt positief is. Voor leerlingen op een schooltype lager dan mavo is de verwachting dat de regressiecoëfficiënt negatief is. Verder is de verwachting dat de regressiecoëfficiënten dezelfde ordening vertonen als het schooltype. Het effect voor bijvoorbeeld vwo-leerlingen zal dan het meest positief zijn en het effect voor ivbo-leerlingen het meest negatief.

Tabel 4.3: Multiniveau analyse. Model met toevoeging van schooltype

	NEDERLANDS	WISKUNDE
Etniciteit (1=allochtoon)	-0,113 (0,026) ***	-0,113 (0,026) ***
Opleidingsniveau vader	0,011 (0,007)	0,011 (0,007)
Opleidingsniveau moeder	0,010 (0,006)	0,010 (0,006)
Aantal ouders (1=twee ouders)	0,004 (0,024)	0,049 (0,020) *
Sekse (1=jongen)	-0,153 (0,019) ***	0,146 (0,016) ***
Géén oudervragenlijst	-0,018 (0,018)	-0,018 (0,018)
Autoritatieve opvoedingsstijl	0,030 (0,006) ***	0,030 (0,006) ***
Praten met ouders	0,006 (0,008)	-0,025 (0,007) ***
Huiswerkcontrole	-0,047 (0,008) ***	-0,020 (0,007) **
Op straat met vrienden	-0,029 (0,006) ***	-0,029 (0,006) ***
Lezen en hobby's	0,058 (0,009) ***	0,010 (0,008)
Bijbaan	-0,010 (0,006)	-0,010 (0,006)
Uitgaan	0,007 (0,009)	-0,028 (0,008) ***
Zelfbeeld	0,040 (0,006) ***	0,040 (0,006) ***
Mening over klasgenoten	0,029 (0,009) **	-0,005 (0,007)
Mening over docenten en regels	0,033 (0,006) ***	0,033 (0,006) ***
Aantal dagen huiswerk	0,016 (0,007) *	0,016 (0,007) *
Non-verbale intelligentie	0,015 (0,010)	0,092 (0,008) ***
Prestatie taal	0,146 (0,015) ***	0,032 (0,011) **
Prestatie informatieverwerking	0,196 (0,013) ***	0,110 (0,011) ***
Prestatie rekenen	0,033 (0,013) *	0,338 (0,013) ***
Sekse * Taal	-0,037 (0,017) *	
Lezen en hobby's * Taal	0,025 (0,007) ***	
Sekse * Rekenen		-0,058 (0,014) ***
Schooltype (Referentiegroep: mavo)		
- ivbo	-0,688 (0,052)} ***	-1,014 (0,046)} ***
- vbo	-0,459 (0,032)} ***	-0,410 (0,029)} ***
- vbo/mavo	-0,014 (0,074)} ***	-0,028 (0,065)} ***
- havo	0,327 (0,033)} ***	0,179 (0,029)} ***
- havo/vwo	0,507 (0,079)} ***	0,343 (0,067)} ***
- vwo	0,543 (0,037)} ***	0,377 (0,033)} ***
Constante	0,141 (0,037)	-0,006 (0,034)

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

In Tabel 4.3 wordt een model weergegeven met de hoofd- en interactie-effecten uit het vorige hoofdstuk plus de effecten van de verschillende schooltypen. Met dit model kan dus worden nagegaan of het schooltype iets toevoegt aan de verklaring van verschillen in prestaties. Uit Tabel 4.3 blijkt dat bijvoorbeeld ivbo-leerlingen gemiddeld genomen ongeveer zeventiende standaarddeviatie lager op Nederlands en één standaarddeviatie lager op wiskunde scoren dan mavo-leerlingen (onder constant houding van de andere kenmerken). Vwo-leerlingen scoren het beste op beide toetsen. Rekening houdend met de overige kenmerken scoren ze gemiddeld meer dan een halve standaarddeviatie hoger op Nederlands en viertiende standaarddeviatie hoger op wiskunde. Vbo/mavo-leerlingen verschillen wat netto prestaties betreft niet van mavo-leerlingen. Ook het onderscheid tussen havo/vwo- en vwo-leerlingen is gering.

Voor de verschillende typen geldt dat het effect voor ivbo-leerlingen negatiever is voor wiskunde dan voor Nederlands. Havo- en vwo-leerlingen scoren significant hoger op Nederlands dan op wiskunde. Uit de *simultane toets* (Goldstein, 1995²) voor de zes dummy's voor schooltype blijkt dat het contrast 66,80 is met zes vrijheidsgraden, wat overeenkomt met een significantieniveau kleiner dan 0,001. Hieruit kan worden opgemaakt dat het schooltype dus een verschillend effect heeft op Nederlands en wiskunde. In Tabel 4.3 zijn effecten die verschillend zijn voor beide vakken met vet aangegeven.

Door de opname van het schooltype zijn de effecten van het *opleidingsniveau van de vader en de moeder* niet meer significant. Deze effecten worden dus gemedieerd door het schooltype. Naarmate ouders hoger zijn opgeleid, presteren hun kinderen beter in leerjaar drie. Wanneer er rekening wordt gehouden met de prestaties uit de brugklas wordt dit effect minder sterk (hoofdstuk 3). Dit effect valt weg door de toevoeging van het schooltype aan het analysemodel. Er is dus geen direct effect meer van sociaal-economische status op prestaties van leerlingen. Het schooltype intermedieert het effect van sociaal milieu op prestaties.



Naarmate de sociaal-economische status hoger is, presteren leerlingen gemiddeld genomen hoger. De correlatie tussen de opleiding van de moeder, een van de indicatoren van sociaal-economische status, en de toetsscore op Nederlands is bijvoorbeeld 0,29. In een analyse met structurele kenmerken (Bijlage 3A, model 5) is het effect van de opleiding van de moeder op de prestaties op Nederlands 0,056. Dit effect is kleiner dan de correlatie, omdat de opleiding van de moeder matig

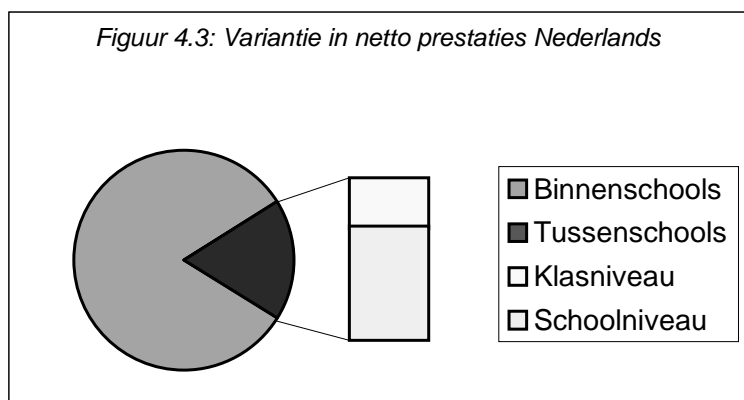
gecorrleerd is met de andere onafhankelijke variabelen, zoals het opleidingsniveau van de vader, het beroepsniveau van de kostwinner en de etniciteit. Vervolgens worden er allerlei culturele kenmerken aan het model toegevoegd, zoals de opvoedingsstijl van ouders of het leefpatroon van leerlingen. Deze variabelen zijn ook gecorrleerd met de opleiding van de moeder en verklaren een deel van het effect. Als vervolgens ook nog de eerdere prestaties van leerlingen worden toegevoegd (Tabel 3.5), dan is het directe effect van de opleiding van de moeder 0,029. Dit effect is klein, maar nog steeds relevant, zoals ik in paragraaf 3.4 naar voren heb gebracht. De vraag is vervolgens waarom heeft de sociaal-economische achtergrond nog steeds een effect op de prestaties van leerlingen. In Tabel 4.3 is het schooltype aan de analyses toegevoegd. Onder constanthouding van overige kenmerken (waaronder de eerdere prestaties) blijkt het schooltype in leerjaar drie een effect te hebben op de prestaties en door de opname van dit effect verdwijnt de invloed van de opleiding van de moeder. De parameterschatting is 0,010 en de standaardfout is 0,006 (Tabel 4.3). Het opleidingsniveau van de moeder is namelijk matig gecorrleerd aan het schooltype, zelfs als er rekening is gehouden met prestaties van leerlingen in leerjaar één.^{xi} Naarmate het sociaal milieu hoger is, komen leerlingen niet alleen op hogere schooltype doordat ze beter presteren, maar waarschijnlijk ook doordat ze vaker hoger kiezen dan hun prestaties aangeven. Bij de overgang van basis- naar voortgezet onderwijs is bekend dat leerlingen uit hogere milieus allereerst een hoger advies krijgen dan hun prestaties aangeven en ten tweede hoger kiezen dan hun advies aangeeft (Dronkers, 1997). Na de brugklas is er op veel scholen nog een keuzemoment en ook dan geldt weer dat leerlingen uit hogere sociale milieus hoger kiezen dan hun prestaties aangeven. Op die hogere schooltypen krijgen ze vervolgens moeilijkere stof, waardoor ze in de loop van het voortgezet onderwijs meer leren dan leerlingen die naar een lagere schooltype gaan. Op de keuzemomenten in het onderwijs ontstaat een belangrijk deel van de ongelijkheid naar sociaal milieu.

Verder valt uit een vergelijking van Tabel 3.5 en Tabel 4.3 op dat ook de effecten van het al dan niet hebben van een *bijbaan* en van het aantal dagen dat een leerling *huiswerk* maakt niet langer significant is in een model met schooltype. Voor het maken van huiswerk geldt dus duidelijk dat dit vooral wordt bepaald door het schooltype van een leerling. Naarmate het schooltype hoger is, maken leerlingen gemiddeld genomen meer dagen in de week huiswerk. Ook geldt dat de effecten van de *eerdere prestaties* minder sterk worden. Uit deze veranderingen in de parameters blijkt dat het schooltype van leerlingen een indicatie is van hun prestatieniveau. De parameters voor eerdere prestaties nemen af. Het schooltype is dus een variabele die met talrijke andere kenmerken samenhangt. Voor onderzoek

naar verschillen in prestaties maakt het daarom sterk uit of er wel of niet rekening wordt gehouden met het schooltype.

Variantiecomponenten en modelfit

Door de opname van het schooltype in de analyse worden de variantiecomponenten beduidend kleiner. Deze componenten bedragen voor Nederlands: op leerlingniveau 0,398; op klasniveau 0,027; op schoolniveau 0,059. De residuele intra-klasse correlatiecoëfficiënt tussen klassen en scholen is $(0,027 + 0,059) / (0,397 + 0,027 + 0,059) = 0,18$ en tussen scholen $0,059 / (0,027 + 0,059) = 0,69$. In Figuur 4.3 is de netto variantie in prestaties Nederlands afgebeeld. De relatieve binnenschoolse variantie bedraagt 82 procent en de relatieve tussenschoolse variantie 18 procent. Deze tussenschoolse variantie betreft voor eenderde variantie tussen klassen en voor tweederde variantie tussen scholen.

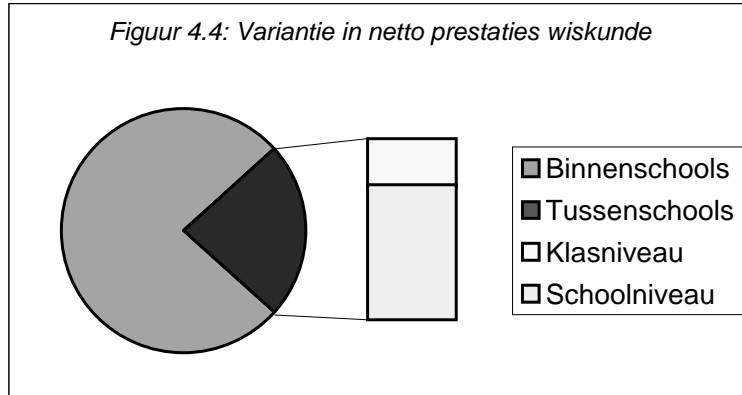


Voor wiskunde bedragen deze componenten op leerlingniveau 0,255; op klasniveau 0,021; op schoolniveau 0,057. Na toevoeging van schooltype bedraagt voor wiskunde de residuele variantie tussen klassen en scholen 23 procent, zie Figuur 4.4. Deze variantie tussen klassen en scholen valt op te delen in 27 procent op klas- en 73 procent op schoolniveau. Voor Nederlands en wiskunde wordt door het schooltype respectievelijk 5 en 7 procent extra variantie op individueel niveau verklaard. Deze gegevens worden in Tabel 4.7 samengevat onder het kopje 'schooltype'.

Met de opname van het schooltype is er op individueel niveau zo uitgebreid mogelijk gecontroleerd. De residuele intra-klasse correlatiecoëfficiënt voor verschillen tussen klassen en scholen is voor Nederlands gedaald van 0,51 voor het lege model tot 0,18 voor het model in Tabel 4.3. Voor wiskunde is er een afname van 0,65 naar 0,23. Als er rekening wordt gehouden met verschillen op indivi-

dueel niveau, dan is de variantie tussen klassen en scholen voor Nederlands 18 en voor wiskunde 23 procent.

Wat de verklaarde variantie betreft wordt op klas- en schoolniveau voor Nederlands 81 en voor wiskunde 88 procent van de variantie gebonden door kenmerken op individueel niveau. Zoals eerder geconstateerd, het overgrote deel van de verschillen tussen scholen is het gevolg van verschillen in leerlingpopulatie.



4.4.2 Random effecten

Om na te gaan of klassen en scholen voor alle leerlingen even goed zijn, kan worden gekeken naar *differentiële effecten*. Vandaar dat ik op klas- en schoolniveau voor alle leerling- en gezinskenmerken uit hoofdstuk 3 ben nagegaan of er sprake is van *random hellingen* en *intercept-helling-covarianties*. Wanneer er sprake is van random hellingen dan is het effect van een bepaald kenmerk niet voor alle leerlingen in een klas of school gelijk. Uiteindelijk bleken er op klasniveau differentiële effecten te zijn voor vijf kenmerken, te weten etniciteit, sekse (en dan alleen voor het effect op wiskunde), het aantal dagen huiswerk, de non-verbale intelligentie (alleen voor wiskunde) en de score op de rekentoets (alleen voor wiskunde).

In hoeverre zijn deze random hellingen van belang? Voor etniciteit geldt dat de gemiddelde helling $-0,114$ (vergelijk Tabel 4.3) en de random helling $0,036$ (Tabel 4.4) is. De gemiddelde helling is het vaste effect van etniciteit op prestaties van leerlingen en de random helling is de variantie van dat effect. Wanneer de variantie $0,036$ is, dan is de standaarddeviatie gelijk aan $\sqrt{0,036} = 0,190$. De gemiddelde helling \forall twee standaarddeviaties loopt van $-0,494$ tot $0,266$. Het verschil tussen deze waarden is ongeveer driekwart standaarddeviatie.

Tabel 4.4: Multiniveau analyse. Random effecten

SCHOOLNIVEAU	PAR. (S.E.)
Variantie Nederlands	0,059 (0,009)
Variantie wiskunde	0,056 (0,008)
Covariantie Nederlands-wiskunde	0,032 (0,007)
KLASNIVEAU	
Variantie Nederlands	0,026 (0,005)
Variantie wiskunde	0,010 (0,003)
Covariantie Nederlands-wiskunde	0,005 (0,003)
Random helling van etniciteit	0,036 (0,015)
Random helling van sekse (wiskunde)	0,009 (0,004)
Random helling van aantal dagen huiswerk	0,004 (0,001)
Random helling van non-verbale intelligentie (wiskunde)	0,005 (0,002)
Random helling van rekenen (wiskunde)	0,007 (0,002)
Intercept-helling covariantie (wiskunde, rekenen)	-0,005 (0,002)
LEERLINGNIVEAU	
Variantie Nederlands	0,393 (0,007)
Variantie wiskunde	0,241 (0,005)
Covariantie Nederlands-wiskunde	0,041 (0,004)

NOOT: Rekening houdend met de vaste effecten uit Tabel 4.3

Het effect van etniciteit op prestaties van leerlingen is niet in alle klassen hetzelfde. In sommige klassen is dit effect sterker en in andere klassen zwakker, wat inhoudt dat voor allochtone leerlingen de ene klas dus gunstiger en de andere klas ongunstiger is. Waarom dit differentiële effect optreedt is vervolgens de vraag. Komt dit bijvoorbeeld door de manier van lesgeven door docenten? Op soortgelijke wijze kan het belang van de andere random hellingen worden bepaald. In Tabel 4.5 heb ik de intervallen weergegeven waarin de effecten van 95 procent van de scholen liggen.

Tabel 4.5 Voorspellingsinterval voor (random) effecten op klasniveau

	GEM.	SD	95% VI
Etniciteit	-0,114	0,190	-0,494; 0,266
Sekse (wiskunde)	0,146	0,094	-0,042; 0,334
Aantal dagen huiswerk	0,015	0,061	-0,107; 0,137
Non-verbale intelligentie (wiskunde)	0,092	0,072	-0,052; 0,236
Rekenen (wiskunde)	0,339	0,083	0,173; 0,505

Behalve random intercepts zijn er op klas- en schoolniveau ook intercept-helling covarianties berekend (Snijders & Bosker, 1999). Wat deze covarianties betreft heb ik allereerst gekeken naar *covarianties tussen het intercept van de afhankelijke variabelen en de helling van de soortgelijke toetsen* uit leerjaar één, dus taal met Nederlands, informatieverwerking met Nederlands, rekenen met wiskunde. Van deze covarianties blijkt alleen die tussen rekenen en wiskunde op klasniveau van belang. Met deze covariantie en de varianties van rekenen en wiskunde kan de correlatie tussen intercept en helling worden berekend. Deze correlatie is: $-0,005 / \sqrt{(0,007 * 0,010)} = -0,60$. Aangezien alle continue variabelen gestandaardiseerd zijn, correspondeert het intercept met de wiskundescore van een leerling met een gemiddelde score op de rekentoets in een klas met een gemiddelde rekentoetsscore. De negatieve correlatie duidt erop dat in klassen met een gemiddeld hoog prestatieniveau (veelal havo- en vwo-klassen) het effect van de rekentoets op de wiskundescore minder sterk is. Het gemiddeld hoge niveau in leerjaar drie van havo- en vwo-klassen komt daarom meer door de goede prestaties van leerlingen die laag scoorden op de rekentoets uit leerjaar één dan door de goede prestaties van leerlingen die hoger scoorden op die toets. Voor (i)vbo-klassen geldt dat het gemiddeld lage niveau in leerjaar drie op wiskunde meer komt doordat de betere leerlingen in het beroepsonderwijs achterblijven bij hun verwachte score dan doordat de zwakkere leerlingen boven hun verwachte score presteren. De verschillen tussen leerlingen in wiskunde gaan dus per schooltype meer op elkaar lijken. In de hogere schooltypen presteren de ‘zwakkere leerlingen’ (met een voor havo- en vwo-leerlingen lage score op de toets uit leerjaar één) boven verwachting op wiskunde. In de lagere schooltypen presteren de ‘betere leerlingen’ beneden verwachting op wiskunde.

Vervolgens heb ik gekeken naar de *covarianties tussen het intercept van de afhankelijke variabelen en de helling van enkele structurele kenmerken*. Van de covarianties van het intercept van Nederlands of wiskunde met de helling van etniciteit, de opleidingsniveaus van vader en moeder en de sekse, blijkt geen van belang.

In totaal zijn er dus vijf random hellingen en één covariantie tussen intercept en helling gevonden. Voor al deze random effecten geldt dat ze differentiëren op klasniveau. Op schoolniveau bleken er geen differentiële effecten te zijn. De kanttekening hierbij is dat er voor dezelfde leerlingkenmerken wel random effecten op schoolniveau kunnen worden gevonden, wanneer de random effecten op klasniveau worden weggelaten. Door de opname van random effecten wordt de modelfit beter, zie Tabel 4.7. De deviance neemt af met 63,4. Bij zes vrijheidsgraden is deze afname significant op een niveau van 0,001. Door de opname van de random effecten veranderen de vaste effecten nauwelijks. Deze effecten worden daarom niet opnieuw weergegeven.

4.4.3 Regressie-effecten tussen klassen of scholen

Bij een analyse met variabelen op één niveau kunnen er talrijke interpretatieproblemen ontstaan. Is een effect van de mening over docenten en schoolregels nu een effect binnen of tussen klassen en scholen? Gaat het bij dit kenmerk om verschillen tussen leerlingen of om verschillen tussen klassen en scholen? Om hier antwoord op te geven kunnen leerling- en gezinskenmerken worden geaggregeerd, waarna die geaggregeerde score als een aparte variabele in de analyses kan worden meegenomen.

Van de kenmerken op individueel niveau uit Tabel 4.3 heb ik op *klas- en schoolniveau* de geaggregeerde score genomen. Voor deze kenmerken geldt dat er in geen van de gevallen sprake is van regressie-effecten tussen klassen en scholen (zie paragraaf 2.3 voor het verschil tussen binnenschoolse en tussenschoolse effecten). Op scholen waar leerlingen bijvoorbeeld gemiddeld genomen positiever zijn over hun docenten en de schoolregels, presteren leerlingen dus niet beter dan op scholen waar leerlingen negatiever zijn. De mening over docenten en regels is een effect op leerlingniveau, dus binnen klassen en scholen. Ook voor de andere kenmerken blijken er geen regressie-effecten tussen klassen en scholen te zijn.

Uit nadere analyses blijkt dat er in een *model zonder schooltype* wel *regressie-effecten tussen klassen* zijn, zie Bijlage 4A. In een dergelijk model zijn er positieve effecten op Nederlands en wiskunde van het gemiddelde opleidingsniveau van de vader (cirkel vijf), de gemiddelde autoritatieve opvoedingsstijl (cirkel vier), het gemiddelde aantal dagen huiswerk, het gemiddelde taalniveau in leerjaar één, en het gemiddelde rekenniveau in leerjaar één (cirkel één). Wanneer er rekening wordt gehouden met het schooltype, verdwijnen deze contexteffecten.^{xii} Deze effecten hangen dus duidelijk samen met dat kenmerk. Naarmate het schooltype hoger is, is het beginniveau ook hoger, maken leerlingen vaker huiswerk, voeden ouders hun kinderen meer autoritatief op en is het opleidingsniveau van de vader hoger.

Het gemiddelde niveau in een klas kan dus een invloed hebben op de prestaties van leerlingen, maar die invloed hangt samen met het schooltype waarin een leerling terecht komt. Wanneer in één analyse leerlingen van verschillende schooltypen worden samengenomen en er gekeken wordt naar regressie-effecten tussen klassen en scholen, moet er worden gecontroleerd voor het schooltype. Wanneer dat niet gebeurt kunnen er effecten van geaggregeerde kenmerken worden gevonden die eigenlijk aangeven in wat voor schooltype een leerling zit. Uit deze analyses komen geen effecten van de samenstelling van een klas of school naar voren.

4.5 Effectieve scholen

In de voorgaande paragraaf heb ik verschillende controles op klas- en schoolniveau uitgevoerd. Uit deze controles blijkt dat de verklaringen van de prestaties van leerlingen aanzienlijk verbeteren, wanneer rekening wordt gehouden met het schooltype van leerlingen. Op klasniveau blijken er verder talrijke differentiële effecten te zijn. Voor verschillende groepen leerlingen maakt het dus uit in welke klas ze zitten. Wanneer er rekening wordt gehouden met het schooltype, dan blijken er geen regressie-effecten tussen klassen of scholen te zijn. Op klas- en schoolniveau kan echter ook worden gedacht aan *vaste effecten* van schoolkenmerken. Ik heb daarom aan het model met vaste en random effecten van leerlingen en gezinskenmerken schoolkenmerken toegevoegd.

Contextuele schoolkenmerken

Het percentage allochtone leerlingen, de locatiegrootte en de denominatie blijken invloed te hebben op de prestaties van leerlingen. In Tabel 4.6 worden deze effecten weergegeven. Aangezien het effect van deze kenmerken niet voor alle schooltypen hetzelfde hoeft te zijn, is ook gekeken of deze variabelen interacteren met het schooltype van leerlingen. Uit deze analyses per schooltype, zie Bijlage 4B en 4C, blijkt bijvoorbeeld dat er voor het effect van het percentage allochtone leerlingen op de toetsscore Nederlands een onderscheid moet worden gemaakt tussen enerzijds leerlingen met een mavo-opleiding of lager en anderzijds havo/vwo-leerlingen.

Voor alle leerlingen geldt dat het percentage allochtone leerlingen een negatief effect heeft op de wiskundeprestaties. Een afwijking van één standaarddeviatie in het percentage allochtone leerlingen, vermindert de wiskundeprestaties met 0,03. Voor leerlingen die op de havo of hoger zitten geldt dat ze niet alleen op wiskunde maar ook op Nederlands slechter presteren naarmate het percentage allochtone leerlingen op school stijgt. Om het effect van havo/vwo-leerlingen te bekijken moeten het hoofd- en interactie-effect worden samengenomen. Een toename van één standaarddeviatie in het percentage allochtone leerlingen, leidt tot een verschil van $-0,085$ ($0,006 - 0,091$) op de toets Nederlands.

Het percentage allochtone leerlingen hangt samen met talrijke andere kenmerken van de leerlingencompositie. De prestaties van leerlingen nemen af naarmate er meer allochtone leerlingen op een school zitten. Zijn die prestaties lager, omdat het percentage allochtone leerlingen een meting is van ongemeten leerlingkenmerken? Of passen docenten op scholen met veel allochtone leerlingen hun lessen aan en verlagen ze hun standaarden? Deze vragen kunnen in dit onderzoek niet worden beantwoord. Het is bovendien ook onduidelijk waarom er voor ivbo-, vbo- en mavo-leerlingen geen effect op Nederlands is. Zijn er voor scholen met

veel allochtone leerlingen verschillen tussen hogere en lagere vormen van voortgezet onderwijs?

Voor leerlingen in het beroepsonderwijs maakt de locatiegrootte (\log_{10}) een verschil uit voor de prestaties op beide vakken. Voor mavo-, havo- en vwo-leerlingen is er geen effect van de locatiegrootte. Een verschil van één standaarddeviatie leidt voor (i)vbo-scholen tot een verschil van 0,070 (-0,002 – 0,068) standaarddeviatie in de toetsscores. Een standaarddeviatie verschil in de locatiegrootte blijkt ongeveer gelijk aan een verdubbeling van het leerlingenaantal, bijvoorbeeld een toename van 300 tot 600 of van 600 tot 1200. Bij een dergelijke verandering dalen de scores van leerlingen in het beroepsonderwijs met eenveertiende standaarddeviatie. De praktische betekenis van deze effecten is groot, omdat ze over vele leerlingen worden uitgespreid (Bosker & Scheerens, 1989).

Er blijkt ook een effect van de denominatie te zijn.^{xiii} Hierbij zijn er geen verschillen naar vak (het contrast is 6,69 bij zes vrijheidsgraden). Wel blijkt er een apart effect te zijn op Nederlands voor gereformeerd-vrijgemaakte mavo-scholen. Vandaar dat het model voor alle schooltypen is uitgebreid met een interactieterm om deze relatie weer te geven.^{xiv} In Tabel 4.6 worden de effecten van de denominatie gegeven.

Wanneer er rekening wordt gehouden met vaste en random effecten van kenmerken op individueel niveau en met de andere contextuele kenmerken, dan hebben algemeen bijzondere en protestants-christelijke scholen de hoogste scores. Scholen van deze denominatie verschillen echter niet significant van openbare scholen, de referentiegroep. De standaardfouten zijn immers (bijna) net zo groot als de parameterschattingen. De laagst scores zijn gemiddeld genomen van interconfessionele scholen. De verschillen tussen denominaties zijn maximaal 0,21 standaarddeviatie.

In eerder onderzoek werden doorgaans vier denominaties onderscheiden: openbaar, protestants-christelijk, rooms-katholiek en overig bijzonder. Reformatorische en gereformeerd-vrijgemaakte scholen werden dan ingedeeld bij protestants-christelijke of overig bijzondere scholen. Algemeen bijzondere en interconfessionele scholen vielen dan beide onder de noemer overig bijzonder. Uit Tabel 4.6 blijkt dat het verfijnen van de meting van denominatie door de opname van zeven in plaats van vier categorieën zinvol is. Algemeen bijzondere scholen scoren wat netto prestaties betreft significant hoger dan interconfessionele scholen. Ook tussen protestants-christelijke, reformatorische en gereformeerd-vrijgemaakte scholen zijn er verschillen.^{xv}

Tabel 4.6: Multiniveau analyse. Spaarzaam model met toevoeging van contextuele en schoolkenmerken

	CONTRAST	O ²	NEDERLANDS	WISKUNDE
CONTEXTUELE KENMERKEN				
% Allochtone leerlingen	0,03	3,60	0,006 (0,014)}	-0,031 (0,015) *
Havo/vwo * % All. lln			-0,091 (0,053)}	
Locatiegrootte (log ₁₀)	-0,03	1,22	-0,002 (0,023)} ***	-0,002 (0,023)} ***
(I)vbo * locatiegrootte	-0,04	1,52	-0,068 (0,019)} ***	-0,068 (0,019)} ***
Denominatie (Referentiegroep: openbaar)				
- algemeen bijzonder	0,12	1,56	0,040 (0,072)} *	0,040 (0,072)} *
- rooms-katholiek	0,01	0,07	-0,027 (0,047)} *	-0,027 (0,047)} *
- interconfessioneel	0,09	0,87	-0,171 (0,074)} *	-0,171 (0,074)} *
- protestants-christelijk	0,03	0,17	0,035 (0,051)} *	0,035 (0,051)} *
- reformatorisch	-0,21	2,23	-0,138 (0,116)} *	-0,138 (0,116)} *
- gereformeerd-vrijgemaakt	-0,01	0,08	0,015 (0,107)} *	0,015 (0,107)} *
- mavo * geref.-vrijgemaakt			0,199 (0,084)} *	
SCHOOLKENMERKEN				
Teamstabiliteit	-0,03	1,35	0,043 (0,019) *	0,043 (0,019) *
Maatregelen bij lesuitval (Referentiegroep: lesuitval)				
- invalregeling	0,16	6,13	0,192 (0,058)} ***	0,031 (0,058)}
- verplichte huiswerkles	0,28	8,19	0,346 (0,087)} ***	0,035 (0,086)}
- havo/vwo * invalregeling			-0,119 (0,061)} ***	
- havo/vwo * huiswerkles			-0,244 (0,109)} ***	
Aard brugperiode (Referentiegroep: dakpansgewijs)				
- homogeen	-0,14	2,15	0,066 (0,052)} ***	0,066 (0,052)} ***
- heterogeen	-0,19	3,33	-0,213 (0,090)} ***	-0,001 (0,092)}
- vwo * heterogeen			0,414 (0,103)} ***	
Mentoraatsysteem (Referentiegroep: smal/verticaal)				
- smal/horizontaal	0,00	0,02	-0,055 (0,044)} *	-0,055 (0,044)} *
- breed/verticaal	0,00	0,03	-0,115 (0,046)} *	-0,115 (0,046)} *
- breed/horizontaal	-0,14	2,70	-0,176 (0,067)} *	-0,176 (0,067)} ***
Docent-leerling ratio	0,01	0,05	0,039 (0,019)} ***	0,039 (0,019)} ***
(I)vbo * Doc.-ll. ratio	-0,05	1,35	-0,089 (0,023)} ***	-0,089 (0,023)} ***

NOOT: Op individueel niveau zijn de kenmerken uit Tabel 4.3 meegenomen.

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Schoolkenmerken

Een aantal schoolkenmerken heeft een effect op de prestaties van leerlingen. Onder constant houding van de leerling- en gezinskenmerken blijkt de teamstabiliteit een zwak positief effect te hebben op de leerprestaties op beide vakken, zie Tabel 4.6. Door één standaarddeviatie verschil in de teamstabiliteit, gaan de prestaties van leerlingen met 0,04 standaarddeviatie vooruit. Wanneer lesuitval wordt opgevangen door een invalregeling of een verplichte huiswerkkles, dan presteren ivbo-, vbo- en mavo-leerlingen op Nederlands een vijfde tot eenderde standaarddeviatie hoger dan wanneer het uur komt te vervallen. Voor havo- en vwo-leerlingen gaat dit effect niet op. Er blijkt namelijk een interactie te zijn tussen de invalregeling en het schooltype.

Van de aard van de brugperiode gaat ook een effect uit op Nederlands. Het blijkt dat leerlingen in heterogene brugklassen het slechtst presteren op Nederlands. Dit effect geldt echter niet voor vwo-leerlingen, zie bijlage 4B. Deze leerlingen presteren niet slechter, wanneer leerlingen in de brugperiode met leerlingen uit meer dan drie schooltypen in een klas zitten. Behalve voor vwo-leerlingen wordt er dus een negatief effect gevonden van heterogeen groeperen op de toetscore Nederlands. Deze bevinding wijkt af van wat in eerder onderzoek is gevonden. Een verschil met dat eerdere onderzoek is echter dat er hier niet alleen een heterogene en homogene, maar ook een dakpansgewijze brugperiode wordt onderscheiden. Leerlingen in dakpansgewijze brugklassen verschillen niet significant in prestaties op beide vakken van leerlingen in een homogene brugklas.^{xvi}

Wat de vormgeving van het mentoraatsysteem betreft valt het op dat hoe uitgebreider dat systeem is opgezet, hoe slechter de prestaties van leerlingen. Alle vormen scoren lager dan een smal-verticaal mentoraatsysteem, de referentiegroep. Voor Nederlands is het verschil tussen het breed-horizontale en het smal-verticale systeem significant. De invloed van een uitgebreid mentoraatsysteem is dus tegen de verwachting in negatief. Dit negatieve effect vraagt om vervolgonderzoek. In dat vervolgonderzoek zou achterhaald moeten worden hoe nuttig mentorlessen eigenlijk zijn. Ontwikkelen leerlingen meer schoolse of algemene vaardigheden tijdens die lessen? Sporen docenten vaker vroegtijdig problemen van leerlingen op? Kunnen docenten die problemen ook oplossen? Verder is het de vraag of oorzaak en gevolg zouden moeten worden omgedraaid: des te slechter de prestaties van leerlingen, des te uitgebreider de opzet van het mentoraatsysteem. Een bezwaar tegen deze redenering is dat het mentoraatsysteem een tamelijk stabiel kenmerk van een school is. Het systeem bestaat al voordat leerlingen de school binnenkomen. Voor het omdraaien van oorzaak en gevolg zou pleiten dat de instroom van scholen verschilt naar gelang het mentoraatsysteem. Vooralsnog zijn daar echter geen duidelijke aanwijzingen voor. Ook zou kunnen worden gekeken

naar effecten van mentoren in plaats van naar effecten van een mentoraatsysteem. Zijn bepaalde mentoren succesvoller dan anderen, en hebben risicoleerlingen de meeste baat bij een goede mentor?

Tot slot, blijkt de docent-leerling ratio geen effect te hebben dat significant is voor alle schooltypen. Voor (i)vbo-leerlingen geldt echter dat ze slechter presteren naarmate deze ratio hoger is. Hier geldt dat niet alleen ivbo- maar ook vbo-leerlingen baat hebben bij een structuur waarin ze niet al te veel verschillende docenten hebben. Uit Tabel 4.6 valt af te leiden dat een toename van één standaarddeviatie in deze ratio leidt tot een verschil van $-0,050$ ($0,039 - 0,089$) in de toetscores op beide vakken.

Geen effect is er van het prestatieniveau als extra toelatingscriterium op de prestaties van leerlingen. In Bijlage 4D valt er weliswaar een positief effect van dit kenmerk te bespeuren op de toetsscore Nederlands. Wanneer het effect van het prestatieniveau als extra criterium op de toetsscore op beide vakken wordt weggelaten, dan verslechtert het model echter niet significant ($\chi^2 = 3.0$, $df = 2$, $p > 0,20$). De variatie in dit kenmerk blijkt ook niet zo groot. Van alle scholen hanteert 78 procent dat extra criterium. Als er wat verfijnder naar het toelatingsbeleid van scholen zou kunnen worden gekeken, dan zouden er misschien verschillen tussen scholen te ontdekken zijn. Zo verschillen zelfstandige gymnasia mogelijk van andere vwo-scholen door een onderscheid in toelatingsbeleid. Vooralsnog zijn er geen aanwijzingen dat een dergelijk effect optreedt.

Naarmate de spreiding in antwoorden tussen docenten op vragen over het schoolklimaat kleiner is en docenten dus meer eenheid in opvattingen vertonen, zijn de netto prestaties op een school niet hoger. Er is dus geen effect van eenheid in visie en doelen op de prestaties van leerlingen. Ook het kenmerk afspraken over orderegels blijkt geen effect te hebben op de prestaties op beide vakken. Uit de Bijlagen 4B, 4C en 4D kan worden opgemaakt dat er apart noch integraal effecten zijn van dit kenmerk.

In Tabel 4.6 zijn alle effecten op de leerprestaties weergegeven. In de tabel is ook aangegeven of effecten voor beide vakken gelijk of verschillend zijn. Uitgaande van een significantieniveau van 0,10 is de kritieke waarde in de chi-kwadraatverdeling 2,71. Vier effecten blijken dan te verschillen naar vak, te weten het percentage allochtone leerlingen, de maatregelen bij lesuitval (voor zowel de invalregeling als de verplichte huiswerkkles) en de heterogene brugperiode.

Tabel 4.7: Multiniveau analyse. Variantie- en covariantiecomponenten en modelfit

	VARIANTIE NEDERLANDS	VARIANTIE WISKUNDE	COVARIANTIE	DEVIANCE)
INDIVIDUEEL *	PAR. (S.E.)	PAR. (S.E.)	PAR. (S.E.)		
Leerlingniveau	0,420 (0,008)	0,270 (0,005)	0,064 (0,005)	22.984,1	
Klasniveau	0,051 (0,007)	0,050 (0,006)	0,030 (0,005)		
Schoolniveau	0,064 (0,011)	0,094 (0,015)	0,050 (0,010)		
SCHOOLTYPE					
Leerlingniveau	0,397 (0,007)	0,255 (0,005)	0,045 (0,004)	22.089,3	894,8
Klasniveau	0,027 (0,005)	0,021 (0,004)	0,008 (0,003)		(df=12)
Schoolniveau	0,059 (0,010)	0,057 (0,009)	0,033 (0,006)		p<0,001
RANDOM EFFECTEN					
Leerlingniveau	0,393 (0,007)	0,241 (0,005)	0,041 (0,004)	22.025,9	63,4
Klasniveau	0,026 (0,005)	0,010 (0,003)	0,005 (0,003)		(df=6)
Schoolniveau	0,059 (0,009)	0,056 (0,008)	0,032 (0,007)		p<0,001
CONTEXTUELE KENMERKEN					
Leerlingniveau	0,393 (0,007)	0,241 (0,005)	0,041 (0,004)	21.965,0	60,9
Klasniveau	0,023 (0,004)	0,010 (0,003)	0,004 (0,003)		(df=20)
Schoolniveau	0,047 (0,008)	0,046 (0,007)	0,025 (0,006)		p<0,001
SCHOOLKENMERKEN					
Leerlingniveau	0,392 (0,007)	0,242 (0,005)	0,041 (0,004)	21.879,3	85,7
Klasniveau	0,023 (0,004)	0,010 (0,003)	0,004 (0,003)		(df=29)
Schoolniveau	0,028 (0,006)	0,040 (0,007)	0,018 (0,005)		p<0,001
SPAARZAAM MODEL					
Leerlingniveau	0,392 (0,007)	0,241 (0,005)	0,041 (0,004)	21.897,8	18,5
Klasniveau	0,023 (0,004)	0,009 (0,003)	0,004 (0,003)		(df=21)
Schoolniveau	0,031 (0,006)	0,043 (0,007)	0,016 (0,005)		p>0,60

* Spaarzaam model met interactie-effecten

In Tabel 4.7 heb ik de variantiecomponenten en de modelfit weergegeven. De resultaten hieruit zijn gedeeltelijk al besproken in de vorige paragraaf. Door de opname van het schooltype (4.4.1) en random effecten (4.4.2) neemt de deviance af. Door opname van de contextuele en de schoolkenmerken wordt de modelfit nog beter. De extra verklaarde variantie door deze kenmerken bedraagt op schoolniveau voor Nederlands 7,3 procent en voor wiskunde 2,9 procent.

In het spaarzame model is de verklaarde variantie voor Nederlands tot 58,0 op leerling-, 87,0 op klas- en 88,6 procent op schoolniveau. Voor wiskunde is de verklaarde variantie dan respectievelijk 74,0, 91,5 en 90,7 procent.

Op klas- en schoolniveau blijkt de structuur van concentrische cirkels geen meerwaarde te hebben. De effecten van contextuele kenmerken kunnen niet of nauwelijks worden verklaard door schoolkenmerken, zie Bijlage 4D. In dit hoofdstuk is dan ook geen extra inzicht verschaft in de relaties tussen deze kenmerken onderling. De aanname dat prestaties van leerlingen meer worden verklaard door schoolkenmerken dan door contextuele kenmerken wordt niet door de analyses ondersteund. Door de toevoeging van contextuele kenmerken gaat de deviance met 60,9 omlaag bij 20 vrijheidsgraden, zie Tabel 4.7. Door de toevoeging van de schoolkenmerken neemt deze maat met 85,7 af bij 29 vrijheidsgraden. Wanneer de afname in deviance wordt afgezet tegen het aantal vrijheidsgraden dan is de verbetering van het model door beide soorten kenmerken vergelijkbaar.

4.6 Besluit

In *bruto prestaties* variëren scholen sterk. De leerlingenpopulatie verschilt immers sterk van school tot school en uit het vorige hoofdstuk blijkt dat kenmerken van leerlingen en gezinnen een sterke invloed op prestaties hebben. Het is dus geen wonder dat er zulke verschillen zijn. In *netto prestaties* zijn die verschillen veel minder groot. Wanneer dan geprobeerd wordt te ontrafelen waarom bepaalde scholen hogere netto prestaties behalen dan andere scholen dan blijkt dat een moeilijke opgave.

Eén van de belangrijkste uitkomsten uit dit hoofdstuk is dat, om de vraag naar verschillen in prestaties te beantwoorden, het van belang is om het *schooltype* in de analyses op te nemen. Het schooltype blijkt een zelfstandig effect te hebben op de leerprestaties en medieert bovendien andere effecten, zoals het effect van het opleidingsniveau van ouders of het aantal dagen waarop door leerlingen huiswerk wordt gemaakt.

Deze bevinding werpt een nieuw licht op onderzoek van Bosker (1990). Bosker heeft onderzocht op welke momenten in de schoolloopbaan ongelijkheid naar sociaal milieu ontstaat. Uit zijn analyses blijkt dat 68 procent van de verschillen naar sociaal-economische status in de voorschoolse periode of tijdens het basisonderwijs ontstaat. De overgang van basisschool naar voortgezet onderwijs heeft betrekking op 14 procent van de verschillen. Tijdens het voortgezet onderwijs ontstaat de resterende 18 procent aan ongelijkheid. Uit dit hoofdstuk blijkt dat die laatste vorm van ongelijkheid voor een belangrijk deel het gevolg is van de schoolkeuze die leerlingen maken na de brugklas. Net als bij de overgang van ba-

sis- naar voortgezet onderwijs kiezen leerlingen uit hogere sociale milieus bij de overgang na de brugklas hoger dan hun prestatieniveau aangeeft. In het voortgezet onderwijs zijn er dus geen directe effecten meer van de sociaal-economische achtergrond op de prestaties van leerlingen. Deze effecten worden volledig gemedieerd door andere leerling- en gezinskenmerken of door het schooltype. Ongelijkheid ontstaat in deze fase dus voor een belangrijk tijdens keuzemomenten.

Een tweede controle die in dit hoofdstuk is uitgevoerd heeft betrekking op *differentiële effecten*. Voor verschillende leerlingen blijkt het uit te maken in welke klas ze zitten. Het effect van etniciteit op prestaties is bijvoorbeeld niet in alle klassen hetzelfde. In sommige klassen is dit effect sterker en in andere klassen zwakker, wat inhoudt dat voor allochtone leerlingen de ene klas gunstiger en de andere klas ongunstiger is. Met multiniveau analyse kunnen dergelijke random effecten worden gemodelleerd. Vervolgens is het dan de vraag of deze random effecten op klasniveau kunnen worden verklaard door andere kenmerken. In dit onderzoek is het niet gelukt om die verklaring te geven, doordat er geen gebruik kon worden gemaakt van klaskenmerken.

Een ander voordeel van multiniveau analyse is dat er gekeken kan worden naar *regressie-effecten tussen klassen en scholen*. Dit soort effecten bleken er niet te zijn. Al zou het percentage allochtone leerlingen wel kunnen worden beschouwd als een tussenschools regressie-effect. Etniciteit blijkt niet alleen op individueel maar ook op schoolniveau een effect te hebben op prestaties van leerlingen. Naarmate er meer allochtone leerlingen op een school zitten presteren leerlingen (allochtoon en autochtoon) slechter. Van de compositie van de school gaat dus een effect uit. Hierbij moet echter worden opgemerkt dat in dit onderzoek het percentage allochtone leerlingen niet de op schoolniveau geaggregeerde score van etniciteit is, maar de door de schooladministratie opgegeven score. Stel, het percentage allochtone leerlingen op een vbo/mavo-school is 30 procent. Deze school heeft vier brugklassen, waarvan er twee in de steekproef komen. In deze twee klassen zitten vijftig leerlingen, waarvan tien allochtoon zijn en veertig autochtoon. De op schoolniveau geaggregeerde score van etniciteit is dus 0,20. Uit dit voorbeeld blijkt dat er een verschil is tussen de opgave door de schooladministratie en de op schoolniveau geaggregeerde score. Deze verschillen blijken in VOCL'93 groot. Van brede scholen zijn in het ene geval alleen maar vbo-klassen genomen en in het andere geval juist vwo-klassen. Om deze vertekening door de selectie van klassen tegen te gaan, wordt voor het meten van het percentage allochtone leerlingen de gegevens van de schooladministratie gebruikt.

Vaste effecten van contextuele en schoolkenmerken blijken er in enkele gevallen te zijn. Een positief effect is er van de teamstabiliteit. Verschillen tussen scholen blijken er ook te zijn naar het mentoraatsysteem. Een uitgebreid mentoraatsys-

teem blijkt negatief samen te hangen met de prestaties beide vakken. Dit verband blijkt tegen de verwachting in negatief. Verder blijken er verschillen te zijn naar denominatie. Gegeven de overige kenmerken, springen bij deze verschillen met name de achterstanden van leerlingen op interconfessionele scholen eruit. De verschillen tussen denominaties zijn overigens beduidend kleiner dan de verschillen tussen openbaar en bijzonder onderwijs in bijvoorbeeld de Verenigde Staten.

De bovengenoemde effecten gelden voor alle schooltypen, maar andere effecten gaan slechts voor een *deel van de schooltypen* of voor slechts *één van de vakken* op. Het percentage allochtone leerlingen heeft voor alle schooltypen een negatief effect op de wiskundeprestaties. Voor havo/vwo-leerlingen geldt dat het effect ook voor Nederlands negatief is. Een ander contextueel kenmerk dat een zwak negatief effect heeft is het effect van de locatiegrootte op de prestaties van leerlingen in het beroepsonderwijs. Ook voor de denominatie-effecten is er sprake van een interactie. Onder constant houding van de overige kenmerken, presteren mavo-leerlingen op gereformeerd-vrijgemaakte scholen beter op Nederlands dan de overige leerlingen.

Verschillen tussen vbo/mavo-scholen blijken er te zijn naar de maatregelen bij lesuitval. Op scholen die gebruik maken van een invalregeling of van verplichte huiswerklessen scoren vbo/mavo-leerlingen beter op Nederlands. De aard van de brugperiode maakt ook uit voor Nederlands. Wanneer leerlingen in de brugperiode heterogeen worden gegroepeerd presteren ze in leerjaar drie slechter op Nederlands. Dit effect blijkt voor leerlingen in alle schooltypen te gelden, met uitzondering van vwo-leerlingen. In de praktijk groeperen scholen leerlingen overigens steeds minder heterogeen. Meer dan de helft van de leerlingen zit in een homogene brugklas. In een heterogene brugklas zit één op de tien leerlingen. De overige leerlingen zitten in een tussenvorm, een dakpanklas.

De docent-leerling ratio heeft alleen een effect voor (i)vbo-leerlingen. Naarmate de ratio groter is, presteren deze leerlingen slechter op beide vakken. Er zijn dus enkele kenmerken die een effect hebben op de prestaties van leerlingen. Deze effecten gaan overwegend dezelfde richting op voor Nederlands en wiskunde, maar verschillen vaak in sterkte.

De multiniveau analyse zijn voor alle schooltypen tegelijkertijd verricht. Door het maken van interacties kunnen echter effecten worden opgenomen die specifiek voor een bepaald schooltype zijn. In het verleden hebben onderzoekers, bijvoorbeeld Van den Eeden, De Jong, Koopman en Roeleveld (1993), betoogd dat een onderzoek naar effectieve scholen per schooltype dient te geschieden. Hun belangrijkste argument is dat het al moeilijk is om op klas- en schoolniveau effecten

te vinden, wanneer er per schooltype wordt geanalyseerd, laat staan dat er een analyse wordt gedraaid voor alle schooltype tegelijkertijd. In dit hoofdstuk heb ik echter laten zien dat een dergelijke analyse wel degelijk kan worden verricht. Effecten die voor een deel van de schooltypen gelden, kunnen met een interactie-term in de analyse worden opgenomen. De meerwaarde van deze integrale analyse is dat het aantal onderzoekseenheden, met name op schoolniveau, niet wordt verkleind. Op die manier kunnen kleine effecten eerder worden opgespoord. Wanneer er afzonderlijk per vak en per schooltype wordt geanalyseerd (Bijlage 4B en 4C), dan heeft teamstabiliteit bijvoorbeeld in zeven van de acht gevallen een positief, maar niet een significant effect. Wanneer het effect van dit kenmerk voor alle schooltypen tegelijkertijd wordt bekeken, blijkt er wel een significant effect te bestaan.

Deze integrale analyses zijn overigens niet afwijkend van gangbaar onderzoek. In landen waar een ongedeelde voortgezet onderwijs is, zoals de Verenigde Staten, vinden analyses voor alle leerlingen tegelijk plaats. In Nederlands onderzoek naar het basisonderwijs is het ook vanzelfsprekend om analyses te verrichten met een steekproef van leerlingen uit het complete spectrum aan scholen, van elitescholen tot alloctonenscholen. Zelfs in het voortgezet onderwijs komen integrale analyses voor. Bosker (1990) heeft bijvoorbeeld gegevens van vbo- tot en met vwo-leerlingen gebruikt in een onderzoek met de positie op de leerjarenladder als afhankelijke variabele. In onderzoek met toetsscores als criteriumvariabele is het echter standaard geworden om analyses per schooltype te verrichten. In dit hoofdstuk is duidelijk gemaakt dat ook voor die analyses een integrale aanpak een meerwaarde heeft.

In onderzoek naar contextuele en schoolkenmerken blijkt de structuur van concentrische cirkels geen meerwaarde te hebben. De effecten van contextuele kenmerken kunnen niet of nauwelijks worden verklaard door schoolkenmerken. In dit hoofdstuk is dan ook geen extra inzicht verschaft in de relaties tussen deze kenmerken onderling. De aanname dat prestaties van leerlingen meer worden verklaard door schoolkenmerken dan door contextuele kenmerken wordt niet door de analyses ondersteund. De toevoeging van beide soorten kenmerken resulteert in een vergelijkbare verbetering van het analysemodel.

In dit hoofdstuk is naar voren gebracht dat de empirische onderbouwing van veel effectieve schoolkenmerken is gebaseerd op buitenlands onderzoek. In tegenstelling tot Amerikaanse scholen verschillen Nederlandse scholen bijvoorbeeld minder in de nadruk op basisvaardigheden. Vandaar dat er ook geen effect van dit kenmerk valt te vinden. Voor het schooleffectiviteitsonderzoek geldt dat de bevindingen uit onderzoek in een bepaald land niet zo maar kunnen worden gene-

raliseerd naar andere landen. Buitenlands onderzoek naar effectieve scholen is uiteraard een goede leidraad voor Nederlands onderzoek op dat terrein. Wanneer het echter een keurslijf wordt, dan laten onderzoekers vragen liggen die specifiek voor ons onderwijssysteem gelden. Het is in dit opzicht opvallend dat er naar de aard van de brugperiode of het mentoraatsysteem nog weinig onderzoek is gedaan.

In de tweede plaats wordt in veel overzichten van effectiviteitsonderzoek geen onderscheid gemaakt naar onderwijsniveau. Er zijn bijvoorbeeld duidelijke verschillen te verwachten in effecten voor scholen voor basis- dan wel voortgezet onderwijs. Waar een effect van de samenwerking tussen school en thuis in het basisonderwijs mogelijk kan worden gevonden, is een dergelijk effect voor het voortgezet onderwijs minder waarschijnlijk.

In de derde plaats wordt in een deel van het effectiviteitsonderzoek onvoldoende gecontroleerd voor kenmerken op individueel niveau. Wanneer rekening wordt gehouden met leerling- en gezinskenmerken, zo blijkt uit dit onderzoek, blijft er weinig variantie op klas- en schoolniveau over. Als vervolgens ook nog wordt gecontroleerd voor het door de leerling bezochte schooltype, dan is er nog minder variantie op hogere niveaus. Verschillen tussen klassen en scholen worden voor meer dan 80 procent verklaard door kenmerken op individueel niveau. Het ontbreken van effecten van sommige schoolkenmerken kan dan ook het gevolg zijn van de beperkte invloed van scholen op prestaties van leerlingen. Volgens Van der Velden (1996) heeft ook onderzoek in het basisonderwijs niet kunnen aantonen dat de zogeheten effectiviteitskenmerken daadwerkelijk effectief zijn.

Toch kan het ontbreken van effecten van schoolkenmerken voor een deel ook komen door het onderzoeksdesign. Schriftelijke vragenlijsten zijn voor het meten van sommige kenmerken op klas- en schoolniveau ongeschikt. Zo kwam in dit hoofdstuk naar voren dat de gegevens uit de docentenvragenlijst weliswaar betrouwbaar, maar niet valide zijn. In het onderstaande wil ik kort ingaan op de vraag in hoeverre school- en docentkenmerken te meten zijn met vragenlijsten. Ook draag ik enkele oplossingen aan om de kwaliteit van de dataverzameling te vergroten.

Voor twee scholen heb ik de data die het CBS, GION en OCTO hebben verzameld vergeleken met gegevens uit veldwerk. Elders heb ik daarover gerapporteerd (Veenstra, 1997). Op grond van deze vergelijking wil ik enkele opmerkingen maken over de opzet van de schriftelijke vragenlijsten van VOCL'93. In de eerste plaats moet het voor directieleden duidelijk zijn of ze de vragenlijst alleen voor hun *locatie* of voor de *gehele school* moeten invullen. Uit de gegevens van VOCL'93 blijkt dat sommige directieleden het voor de locatie en andere het voor de school invullen.

Hierdoor kunnen de antwoorden op sommige vragen niet met elkaar worden vergeleken. Doorgaans is het denk ik het beste om een directielid vragen te stellen over één locatie, omdat locaties sterk van elkaar kunnen verschillen. Antwoorden die voor de hele school gelden zijn dan niet zinnig. Sommige gegevens, zoals de schoolgrootte of de omgang tussen de schooldirectie en docenten, zouden echter wel moeten worden verzameld.

Ten tweede kunnen verscheidene gegevens die nu via de vragenlijsten worden verzameld uit informatiebrochures, jaarverslagen of de toekomstige schoolgidsen worden gehaald. Dergelijk *schriftelijk materiaal* kan bij de scholen worden opgevraagd. Op die manier kunnen de vragenlijsten voor met name de directieleden wat minder omvangrijk worden. Ook kan op deze manier worden gezorgd dat er geen ontbrekende gegevens zijn voor bijvoorbeeld de locatiegrootte, het aantal lesuren of de examenresultaten. Wanneer dergelijke gegevens jaarlijks worden bijgehouden komen storende fouten, zoals een verkeerde denominatie, waarschijnlijk minder vaak voor.

In de derde plaats merk ik op dat het voor de meeste gegevens over scholen en docenten nodig is om *herhaaldelijk metingen* te verrichten. Uit de vergelijking (Veenstra, 1997) blijkt bijvoorbeeld dat de ene school het nascholingsbeleid recent heeft gewijzigd en dat de aard van brugperiode op de andere school van dakpansgewijs is omgezet in heterogeen. Om een goed overzicht van de praktijk in het onderwijs te geven, is het dus nodig om dergelijke informatie om de paar jaar op te vragen. Een groot deel van de gegevens zal echter niet elke keer veranderen. Vandaar dat het misschien mogelijk is om alleen de eerste keer een uitgebreide vragenlijst naar de directie te sturen. Bij latere metingen kan de vragenlijst wellicht worden ingekort door alleen die vragen op te nemen die betrekking hebben op nieuwe ontwikkelingen. Wat de overige gegevens betreft zou het raadzaam kunnen zijn om aan de directie een uitdraai te geven van hun antwoorden op het eerste meetmoment met daarbij de vraag om na te gaan of bepaalde antwoorden inmiddels achterhaald zijn en dus moeten worden veranderd.

Ook voor de prestatiemeting bij leerlingen is een herhaaldelijke meting van belang. Leerlingen in het voortgezet onderwijs wisselen vaak jaarlijks van docent. Het beste onderzoeksdesign is daarom waarschijnlijk een met een prestatiemeting aan het begin en aan het einde van hetzelfde schooljaar. Het ideaal zou dan zijn om deze procedure enkele jaren achtereenvolgend te voeren. Op die manier kan worden nagegaan in hoeverre een leerling vooruit gaat in een bepaald leerjaar (en bij een bepaalde docent) en over verschillende leerjaren (en bij één of meer docenten).

Ten vierde is het nodig om de opzet van de vragenlijst voor docenten te veranderen. De huidige vragenlijsten zijn *te omvangrijk*. Verscheidene docenten hadden

daar geen goed woord voor over en maakten opmerkingen als 'De leukheid was omgekeerd evenredig met de tijd', 'Zonde van m'n tijd' of 'Ik werd steeds agressiever door de grote hoeveelheid vragen! Dus ik stop er nu ook mee!'. In paragraaf 4.3 bleken de vragen waarbij van docenten wordt verlangd om percentages of frequenties te schatten van hun activiteiten, niet valide te zijn. Docenten schudden deze gegevens, zo was mijn indruk tijdens het kwalitatieve onderzoek (Veenstra, 1997), zo uit de mouw. Ze geven altijd op dat ze tijd besteden aan het uitleggen van nieuwe stof, maar in de praktijk blijkt dat een aanzienlijk deel van de docenten een les als volgt opbouwt: ongeveer vijf minuten verloren tijd doordat leerlingen van lokaal moeten wisselen, vijftien minuten huiswerk bespreken, twintig minuten opgaven of huiswerk maken en vijf minuten orde houden, huiswerk opgeven en andere praktische zaken. Het uitleggen van nieuwe stof komt lang niet in elke les voor. Het lijkt er dan ook op rapportages die docenten over zichzelf geven twijfelachtig zijn. Voor het achterhalen van kenmerken van docenten zal moeten worden gekozen voor andere methoden van dataverzameling. Om dergelijke gegevens te achterhalen kan bijvoorbeeld gebruik worden gemaakt van observaties of logboeken, zie bijvoorbeeld De Jong en Westerhof (1998).

Een alternatief voor observaties of logboeken is mogelijk het ondervragen van leerlingen over een docent om zo het gemiddelde oordeel van leerlingen te bepalen (Brekelmans, 1989). Scheerens en Bosker (1997, p. 205) halen een studie van Grisay naar het onderwijs in Frankrijk aan, waar succesvol gebruik is gemaakt van de oordelen van leerlingen over een docent. "The fact that pupil judgments about likely effectiveness-enhancing conditions showed most of the significant process-output associations raises an interesting methodological point. Pupil judgments and observations on school organizational functioning and teaching can be seen as having two advantages: avoiding self-reports by teachers and head teachers, and measuring these process conditions at the level where the impact on actual performance is likely to be greatest." In tegenstelling tot experts of observatoren kunnen leerlingen hun mening baseren op vele lessen. Bovendien is de mening van leerlingen over docenten misschien ook wel relevanter. Docenten kunnen vinden dat ze bijvoorbeeld duidelijk uitleggen, maar die uitleg is pas echt duidelijk als leerlingen dat ook zo ervaren. In onderwijsonderzoek worden oordelen van leerlingen over docenten nog niet zo vaak gebruikt. De informatie die zo verkregen wordt is naar het schijnt echter geschikter dan die uit een vragenlijst voor docenten.

Een andere oplossing voor de problemen met de docentenvragenlijst is om kortere vragenlijsten bij meer docenten af te nemen. Op dit moment worden er alleen vragenlijsten afgenomen bij docenten Nederlands en wiskunde. Voor die vakken maken leerlingen prestatietoetsen en dus ligt het voor de hand om na te gaan of het

onderwijsaanbod in deze vakken en de *wijze van lesgeven* van de betreffende docenten van invloed is op de vooruitgang in prestaties van leerlingen. Vragen over het *functioneren van de school* kunnen echter ook aan andere docenten worden gesteld. Bij die vragen gaat het er niet zo zeer om wat de betreffende docent ervan vindt, maar om wat verschillende docenten gezamenlijk vinden en om de mate waarin hun antwoorden van elkaar verschillen.

Ten vijfde moet het mogelijk zijn om de vragenlijsten *computergestuurd* af te nemen. Tijdens het veldwerk bleek dat docenten het leuk vinden om op die manier een vragenlijst in te vullen. Deze manier van enquêteren heeft verscheidene voordelen. Het doorverwijzen in de vragenlijst kan automatisch gebeuren, waardoor hiermee geen fouten kunnen worden gemaakt. De vragenlijst kan zo ook sneller worden afgenomen. Verder zullen er minder ontbrekende gegevens voorkomen, omdat er pas een nieuwe vraag op het scherm verschijnt, wanneer de vorige vraag is beantwoord. Computergestuurd interviewen komt de laatste jaren steeds vaker voor. Het is goed om na te gaan hoe die methode ook in onderwijsonderzoek kan worden gebruikt.

Eindnoten

ⁱ Een ander kenmerk dat als contextueel kan worden beschouwd is de hechtheid van de gemeenschap rondom de school. Wanneer er rekening wordt gehouden met effecten op individueel niveau zijn er geen aanwijzingen dat dit kenmerk van invloed is op prestatieverschillen tussen scholen (Veenstra & Dijkstra, 1998). Verder kan ook aan de bestuurlijke inrichting van scholen worden gedacht (Chubb & Moe, 1990; Hofman, 1993).

ⁱⁱ Voor het basisonderwijs stellen Jungbluth, Peetsma en Roeleveld (1996) dat er in Nederland sprake is van 'standenonderwijs'. Er zijn namelijk elitescholen (met veel autochtone leerlingen uit gezinnen waar de ouders hoger onderwijs hebben gevolgd), burgerscholen (met veel autochtone leerlingen uit gezinnen waar de ouders lager of middelbaar beroepsonderwijs hebben gevolgd) en allochtonenscholen (met veel allochtone leerlingen uit gezinnen waar de ouders hoogstens lager beroepsonderwijs hebben gevolgd). Uit deze laatste categorie distilleren zij nog een apart type waarop voornamelijk Marokkaanse en Turkse leerlingen zitten.

ⁱⁱⁱ Een aanzet hiertoe is Heyl (1996). Heyl beschrijft de structuur, de intensiteit en de invloed van informele contacten tussen docenten in kleine mavo-scholen. Met prestaties van leerlingen worden deze informele contacten niet in verband gebracht.

^{iv} Net als Oakes (1985) controleren zij onvoldoende voor andere kenmerken. Zie Peschar (1994) voor een bespreking van onderzoek naar de klassensamenstelling.

^v Een kenmerk dat ook in deze cirkel zou passen is hoge verwachtingen van docenten over prestaties van leerlingen. Over de invloed van deze verwachtingen op de daadwerkelijke resultaten is vaak gerapporteerd (Brophy & Good, 1986; Edmonds, 1979; Madaus, Kellaghan, Rakow, & King, 1979; Rutter e.a., 1979; Wilson & Corcoran, 1988). Aan de basis hiervan ligt het onderzoek van Rosenthal en Jacobson (1968) op een basisschool met voornamelijk leerlingen uit lagere sociale milieus. De verwachtingen die een docent over een leerling heeft beïnvloedt, zo vermoeden zij, de houding van de docent tegenover de leerling. De verwachtingen werken op verscheidene manieren

door in het gedrag van docenten tegenover leerlingen, bijvoorbeeld door positieve of negatieve gevoelens, de hoeveelheid informatie en terugkoppeling die zij geven en het aantal keren dat ze een vraag stellen. Volgens Rosenthal en Jacobson is het cruciale hieraan dat docenten die hoge verwachtingen over hun leerlingen koesteren en dat ook naar alle leerlingen uitdragen, betere schoolprestaties bereiken dan docenten met lage verwachtingen. Als een docent zich eenmaal een negatief beeld van leerlingen heeft gevormd, is hij geneigd om dergelijke leerlingen minder aandacht te geven en minder moeilijke taken op te geven. Dit principe noemen zij het Pygmalion-effect en het effect dat dit veroorzaakt staat bekend als een zichzelf waarmakende voorspelling (self-fulfilling prophecy), een voorspelling die een reeks gebeurtenissen beïnvloedt waardoor de aanvankelijke verwachting uitkomt. Een ander kenmerk dat in deze cirkel zou passen is positieve bekrachtiging (Walberg, 1984). Door op antwoorden of huiswerk van leerlingen positief te reageren is het volgens Walberg voor een docent mogelijk om zijn betrokkenheid bij leerlingen te laten blijken, een prettige sfeer in de klas te scheppen, gewenst gedrag te stimuleren en ongewenst gedrag bij te sturen. Leerlingen gaan zich door deze bekrachtiging competentier voelen. Naar hoge verwachtingen en positieve bekrachtiging kan met gegevens uit VOCL'93 geen onderzoek worden gedaan.

^{vi} Enigszins vergelijkbaar is de instructiemethode van Bloom die bekend staat als mastery learning (Block & Burns, 1976). Ook bij deze methode wordt gebruik gemaakt van duidelijke doelen, kleine eenheden in het curriculum, een weloverwogen volgorde in de leerstof en remediale hulp voor lacunes in de kennis of vaardigheden van leerlingen. Het belangrijkste verschil met gestructureerd onderwijs is volgens Bloom dat mastery learning het behalen van een standaardniveau voor alle leerlingen beoogt. Elke persoon kan een bepaald niveau bereiken, zo meent Bloom, mits de condities waaronder hij iets moet leren geschikt zijn. Voor het effect van deze methode is het empirisch houvast stevig.

^{vii} In extra analyses blijken deze verwachtingen te worden bevestigd. Scholen in Nederland blijken niet of nauwelijks te verschillen in de mate waarin ze *basisvaardigheden benadrukken* (geoperationaliseerd als het aantal uren Nederlands en wiskunde dat leerlingen per week krijgen) en als er al verschillen zijn, vertaalt dat zich niet in de prestaties van leerlingen. Van het *professioneel en onderwijskundig leiderschap* gaat evenmin een effect uit op prestaties Nederlands en wiskunde. Net als voor effecten van docentenkenmerken, zie paragraaf 4.3, geldt dat VOCL'93 waarschijnlijk niet zo geschikt is om deze effecten na te gaan. Met VOCL'93 kan alleen worden gekeken naar de tijdsindeling van de directie, waarbij vier categorieën worden onderscheiden: (1) onderwijsinhoudelijk, (2) administratief en financieel; (3) personeel; (4) interne en externe contacten (zie Bijlage 6 voor de oorspronkelijke vraag). In hoeverre de schoolleiding de structuur van en de relaties binnen de school bewaakt en tijd en energie steekt in activiteiten om de school te verbeteren, valt waarschijnlijk moeilijk te achterhalen met een schriftelijke vragenlijst. De *bewaking van leervorderingen* heeft ook geen effect. Over een leerlingvolgsysteem beschikt 87 procent van de scholen niet, 12 procent gedeeltelijk en 1 procent wel. Op dit kenmerk is er dus weinig variatie tussen scholen. Van de *samenwerking tussen school en thuis* gaat evenmin een effect uit. Ook voor dit kenmerk geldt dat er weinig variatie tussen scholen is. De band tussen school en ouders is in het voortgezet onderwijs formeel. Scholen organiseren voor ouders jaarlijks enkele ouder- en thema-avonden. De opkomst op die avonden is volgens de schooldirecties overwegend goed en verschilt tussen scholen niet altijd te sterk. Voor zover er verschillen zijn heeft dat geen invloed op de prestaties van leerlingen.

^{viii} Het effect van het percentage allochtone leerlingen hoeft niet lineair te zijn. Om na te gaan of er een ander verband is, heb ik dit kenmerk gecategoriseerd in: (1) minder dan 6 procent; (2) 6 tot

en met 20 procent; (3) 21 tot en met 40 procent; (4) meer dan 40 procent. Vervolgens heb ik deze categorieën met drie dummy's gerepresenteerd. Het blijkt dat het gebruik van dummy's geen beter model oplevert.

^{ix} De aard van de brugperiode hangt sterk samen met de duur van de brugperiode. In homogene brugklassen is de duur nul jaar, in heterogene brugklassen doorgaans één jaar en in dakpansgewijze brugklassen overwegend twee jaar. Als aparte variabele of in combinatie met de aard van de brugperiode, zo blijkt uit nadere analyses, voegt de brugduur niets toe aan de multiniveau analyse.

^x Willms (1985) laat zien dat schooleffecten te groot uitvallen, wanneer te weinig wordt gecontroleerd.

^{xi} De partiële correlatie tussen de opleiding van de moeder en het schooltype (waarbij gedaan is alsof het een metrische variabele is met een schaal van 0 voor ivbo tot 8 voor vwo) onder constanthouding van de scores op de prestatietoetsen taal, informatieverwerking en rekenen is 0,19.

^{xii} In plaats van op klasniveau kunnen dezelfde vijf leerling- en gezinskenmerken ook op schoolniveau worden geaggregeerd. Er blijken dan soortgelijke regressie-effecten tussen scholen te zijn. Ook deze effecten verdwijnen zodra het schooltype aan de analyses wordt toegevoegd.

^{xiii} Wanneer de denominatie wordt opgenomen, gaat de deviance omlaag met 18,1 bij zeven vrijheidsgraden ($p < 0,05$).

^{xiv} Dit effect was niet voorzien. Het is daarom verstandig om met andere datasets na te gaan of gereformeerd-vrijgemaakte mavo-scholen een aparte positie innemen.

^{xv} In Bijlage 4B valt op dat er negatieve, significante effecten zijn van reformatorisch onderwijs op de prestaties Nederlands van vbo- en havo-leerlingen. In een analyse met alle schooltypen valt dit effect echter weg.

^{xvi} De aard van de brugperiode is een variabele op schoolniveau. Met de gegevens van VOCL'93 hanteren Van der Werf, Lubbers en Kuyper (1999) een soortgelijke variabele op individueel niveau: het door leerlingen gevolgde *klastype*. Rekening houdend met enkele structurele kenmerken (sociaal-economische status en etniciteit), culturele kenmerken (prestatiemotivatie, schoolbeleving en eerdere prestaties) en het door leerlingen gevolgde schooltype, laten zij zien dat leerlingen die in een homogene brugklas hebben gezeten significant beter presteren op Nederlands dan leerlingen in een dakpansgewijze of heterogene brugklas. Voor wiskunde is er geen effect van het klastype. Daarentegen is het voor de onderwijspositie in leerjaar vijf, uitgedrukt in een score op de leerjarenladder, juist relatief minder gunstig om in een homogene brugklas te hebben gezeten. Van der Werf, Lubbers en Kuyper (1999) komen tot de slotsom dat een brugklas waarin een type is vertegenwoordigd dat hoger ligt dan het advies relatief gunstig is voor de onderwijspositie in het vijfde leerjaar. Hierbij geldt als kanttekening dat deze auteurs geen random effect hebben opgenomen voor de brugklas waarin de leerling zat. Zij onderscheiden ten onrechte alleen een individueel en schoolniveau.

5 | Vooruitgang in prestaties

5.1 Inleiding

Welke leerlingen *raken* achterop of *gaan* vooruit in het onderwijs? In hoeverre hebben kenmerken op leerling-, klas- en schoolniveau invloed op de *leervorderingen*? Deze vragen kunnen niet worden beantwoord met de gegevens uit hoofdstuk 3 en 4. In die hoofdstukken staat de vraag centraal in hoeverre de prestaties van leerlingen op één bepaald tijdstip kunnen worden verklaard uit allerlei kenmerken van leerlingen, gezinnen en scholen. Om de netto prestaties van leerlingen te bepalen wordt hierbij rekening gehouden met de prestaties van leerlingen op een eerder tijdstip. In het model dat ik in de vorige hoofdstukken heb gebruikt, zijn de toetsscores in leerjaar drie de afhankelijke variabelen en behoren de eerdere prestaties in leerjaar één tot de onafhankelijke variabelen. Een dergelijk model is standaard in het schooleffectiviteitsonderzoek en staat bekend als het covariantie-analytisch model.

Met het covariantie-analytisch model kunnen prestaties van leerlingen op een bepaald tijdstip worden verklaard. In deze verklaring gaat het om de vraag welke leerlingen een achterstand of voorsprong hebben. Vervolgens kan dan worden nagegaan welke kenmerken invloed hebben op de leerprestaties. Met dit model wordt naar invloeden op leerprestaties en niet naar invloeden op leervorderingen gekeken. Wanneer een onderzoek op dit model is gebaseerd, kan dan ook niet over effecten op vooruitgang in prestaties worden gesproken.

Om naar vooruitgang in prestaties te kijken, moet echter worden gekozen voor een variantie-analytisch model van leerwinst. Met dit model kunnen de vragen aan het begin van deze paragraaf worden beantwoord. Het covariantie-analytisch model wordt door onderzoekers veel vaker gebruik dan een analyse van verschillen in leerwinst. Die populariteit berust waarschijnlijk te veel op traditie, want voor onderwijsonderzoekers zijn niet alleen vragen naar leerprestaties maar ook vragen naar leervorderingen van belang. In dit hoofdstuk zullen beide methoden met elkaar worden vergeleken. Op die manier zal duidelijk worden dat de methoden elkaar aanvullen en dat afhankelijk van het soort vragen dat een onderzoeker stelt, een passend analytisch model moet worden gekozen (Van den Bergh & Kuhlemeier, 1997).

In dit hoofdstuk staat een multivariaat variantie-analytisch model van verschillen in (relatieve) leerwinst centraal. De vragen die aan bod komen zijn:

In hoeverre en waarom zijn kenmerken op individueel, klas- en schoolniveau van invloed op de vorderingen (verschillen in relatieve leerwinst) van leerlingen tussen het eerste en het derde leerjaar van het voortgezet onderwijs?

De afhankelijke variabelen hierbij zijn de verschillen tussen de gestandaardiseerde prestaties in leerjaar drie en leerjaar één voor Nederlands en wiskunde. In paragraaf 5.2 wordt uitgelegd wat verschillen zijn. In die paragraaf bespreek ik ook het model met deze criteriumvariabele. Vervolgens besteed ik aandacht aan de resultaten op zowel leerlingniveau (5.3) als klas- en schoolniveau (5.4). In paragraaf 5.5 vergelijk ik deze resultaten met die uit de twee voorgaande hoofdstukken.

In dit hoofdstuk maak ik opnieuw gebruik van de structuur van concentrische cirkels. Op individueel niveau bestaat de structuur uit de zes cirkels die in hoofdstuk 3 zijn besproken. Op klas- en schoolniveau worden net als in hoofdstuk 4 contextuele en schoolkenmerken meegenomen in de analyses. De analyses zijn gedaan voor dezelfde steekproef van 7.000 leerlingen binnen 450 klassen binnen 150 scholen. Zie de vorige hoofdstukken voor een beschrijving van deze steekproef.

5.2 Leerwinst en verschillen

In deze paragraaf ga ik allereerst in op de vraag hoe in dit onderzoek een vorm van leerwinst kan worden bepaald. Daartoe wordt gebruik gemaakt van verschillen. Vervolgens bespreek ik de verschillen en overeenkomsten tussen een co-variantie-analytisch model en een variantie-analytisch model van leerwinst.

Leerwinst

Naar prestaties van leerlingen is veelvuldig onderzoek gedaan. Veranderingen in prestaties krijgen echter veel minder vaak aandacht van onderwijsonderzoekers. Om veranderingen in prestaties te bepalen, zijn er verschillende mogelijkheden. Wanneer de toets op beide meetmomenten identiek is, dan kan de verandering worden bepaald door de score op de beginmeting van de score op eindmeting af te trekken. Deze verandering in prestaties wordt meestal aangeduid met de term *leerwinst*. Wanneer de prestaties op de eindmeting hoger zijn dan die op de beginmeting is er sprake van positieve leerwinst. Een leerling die veel leerwinst boekt is een leerling die op de eindmeting veel hoger scoort dan op de beginmeting. Mochten de prestaties op de eindmeting lager zijn dan die op de beginme-

ting, dan is er sprake van negatieve leerwinst. Aan het herhaaldelijk gebruiken van een zelfde toetsen kleven enige bezwaren. Positieve leerwinst kan komen doordat leerlingen bij de eindmeting meer weten dan bij beginmeting. Het kan echter ook zo zijn dat leerlingen bij de tweede afname er voordeel van hebben dat ze de toets al eerder hebben gemaakt. Een tweede bezwaar is dat toetsen doorgaans bedoeld zijn voor leerlingen uit een bepaald leerjaar. Het heeft daarom geen zin om dezelfde toets te gebruiken voor leerlingen uit leerjaar één en drie.

Het ideaal is daarom dat leerlingen op twee tijdstippen een toets maken die gedeeltelijk in plaats van geheel overlapt. Bij gedeeltelijke overlap kunnen de toetsen met het Rasch/OPLM-model (Verhelst, Glas & Verstralen, 1995) worden gelijkgeschakeld, zoals in VOCL'93 voor de twee wiskundetoetsen uit leerjaar drie is gedaan (zie bijvoorbeeld: Lee & Smith, 1997). Tussen de toetsen Nederlands en wiskunde uit leerjaar één en drie bestaat echter geen overlap. Het lijkt dus niet mogelijk om de leerwinst te bepalen. Een alternatief voor deze benaderingen is om de scores op beide meetmomenten te standaardiseren om vervolgens alsnog een verschilscore te berekenen door de beginmeting van de eindmeting af te trekken (Fennessey & Salganik, 1983; Kenny, 1975; Kuyper & Veenstra, 1995; Willett, 1988).

Verschilcores

Door de scores op beide toetsen te standaardiseren en per leerling het verschil tussen twee *z-scores* te berekenen, zijn verschilcores berekend voor Nederlands en wiskunde. De toets Nederlands in leerjaar drie is een toets voor tekstbegrip. In leerjaar één ontbrak een aparte toets op dit domein. Tekstbegrip was echter wel een onderdeel in zowel de taal- als de informatieverwerkingstoets. De items voor tekstbegrip uit deze toetsen zijn daarom samengenomen. Deze twaalf items vormen te zamen een schaal voor tekstbegrip, met een betrouwbaarheid van 0,68. Het gemiddelde op deze schaal is 7,6 (standaarddeviatie: 2,6). De verschilcores voor tekstbegrip Nederlands en wiskunde zijn bepaald door van de toetsscore in leerjaar drie de toetsscore in leerjaar één af te trekken. Een beschrijving van deze toetsen heb ik in paragraaf 3.3 gegeven. De correlatie tussen de scores op de toetsen uit leerjaar één en drie is 0,55 voor Nederlands en 0,74 voor wiskunde.

Wanneer de toetsen uit beide leerjaren met elkaar worden vergeleken, dan valt op dat het domein hetzelfde is (tekstbegrip en wiskunde), maar dat de toetsen verschillen in moeilijkheidsgraad. Bij het gebruik van dergelijke verschilcores moet eigenlijk van *relatieve leerwinst* of van verschillen in *relatieve posities* worden gesproken. Absolute leerwinst valt immers alleen te berekenen als de toetsen geheel of gedeeltelijk overlappen. Hier wordt alleen maar naar het verschil

in de relatieve positie van leerlingen op twee tijdstippen gekeken. Het gemiddelde verschil tussen de twee tijdstippen is gelijk aan nul.

In Tabel 5.1 worden de verschillen weergegeven. De verschillen Nederlands (tekstbegrip) voor vwo-leerlingen is bijvoorbeeld gemiddeld genomen 0,18. In het derde leerjaar scoorden vwo-leerlingen gemiddeld 1,06 standaarddeviatie boven het algehele gemiddelde op de toets Nederlands. Twee jaar eerder scoorden deze leerlingen 0,88 standaarddeviatie boven het gemiddelde. Het verschil tussen deze twee scores is $1,06 - 0,88 = 0,18$. Er is geen overduidelijk patroon in de relatie tussen schooltype en gemiddelde verschillen. Ivbo-leerlingen gaan gemiddeld genomen vooruit op Nederlands, maar raken sterk achterop in wiskunde. Behoorlijk vooruit in Nederlands gaan havo/vwo- en vwo-leerlingen. Voor wiskunde valt de vooruitgang van vbo/mavo-leerlingen op.

Tabel 5.1: Verschillen in relatieve positie naar schooltype

NEDERLANDS	GEM.	SD	MIN.	MAX.	N
IVBO	0,14	0,89	-2,73	2,91	416
VBO	-0,12	1,03	-3,55	2,91	1.899
VBO/MAVO	-0,09	1,08	-3,14	2,27	188
MAVO (incl. MAVO/HAVO)	-0,05	1,05	-3,60	4,13	2.092
HAVO	0,05	0,96	-3,05	3,18	806
HAVO/VWO	0,29	0,93	-2,33	2,90	254
VWO	0,18	0,85	-2,33	2,67	938
Totaal	0,00	1,00	-3,60	4,13	6.593
WISKUNDE	GEM.	SD	MIN.	MAX.	N
IVBO	-0,41	1,17	-4,75	2,64	391
VBO	-0,03	1,05	-5,36	3,17	1.712
VBO/MAVO	0,43	0,91	-3,38	2,81	167
MAVO (incl. MAVO/HAVO)	0,09	0,98	-3,68	3,98	1.722
HAVO	-0,05	0,92	-3,04	3,76	753
HAVO/VWO	0,05	0,91	-2,96	2,92	257
VWO	0,01	0,88	-3,05	3,47	840
Totaal	0,00	1,00	-5,36	3,98	5.842

Covariantie-analyse versus verschillen in relatieve posities

Het covariantie-analytisch model en het variantie-analytisch model van leerwinst zijn verscheidene keren met elkaar vergeleken (Allison, 1990; Van den Bergh & Kuhlemeier, 1997; Reichardt, 1979). In het onderstaande schema zijn enkele verschillen tussen beide methoden weergegeven. Waar in de covariantie-benadering een *toestand* op een bepaald moment centraal staat, is het kenmerkende begrip in een analyse van verschillen een *verandering* tussen twee momenten. Naar onderwijsonderzoek vertaald komen deze algemene constructen neer op een analyse naar leerprestaties dan wel leervorderingen. In de covariantie-benadering gaat het om wat een leerling op een bepaald tijdstip weet. In de verschillen-benadering gaat het om wat een leerling op een bepaald tijdstip weet en niet wist op een eerder tijdstip. Het gaat respectievelijk om de vraag waarom bepaalde leerlingen een voorsprong of achterstand hebben of waarom bepaalde leerlingen vooruit of achteruit gaan.

<i>Covariantie-benadering</i>	<i>Analytisch model</i>	<i>Verschilscore-benadering</i>
toestand	algemeen construct	verandering
leerprestaties	specifiek construct	leervorderingen
twee of meer	meetmomenten	twee of meer
toetsscore op t2	afhankelijke variabele	verschilscore (t2-t1)
toetsscore op t1	covariaat	-
voorsprong hebben	positief effect	vooruit gaan
achterstand hebben	negatief effect	achteruit gaan

Voor beide methoden zijn longitudinale gegevens nodig. Er zijn dus altijd twee of meer meetmomenten. In een analyse van verschillen in relatieve posities wordt de verschillen berekend tussen begin- en eindmeting. Deze verschillen is de afhankelijke variabele in een onderzoek naar leervorderingen. Aangezien het om het verschil in scores op *twee tijdstippen* gaat, heeft de verschillen betrekking op verandering. In de covariantie-benadering is de eindmeting de afhankelijke variabele. De afhankelijke variabele is dus de score op *één tijdstip*. Vandaar dat in deze benadering betrekking heeft op een toestand op een bepaald moment. De factor tijd zit in het covariantie-analytisch model niet in de afhankelijke variabele. Door de opname van de beginscore als covariaat worden de scores op de eindmeting gezuiverd van de invloed van de beginmeting. In beide modellen kan vervolgens de invloed van verschillende onafhankelijke variabelen op de criteriumvariabele worden nagegaan, zie het voorbeeld op pagina 143.

Om de vragen naar verschillen in leerprestaties te beantwoorden, kan gebruik worden gemaakt van de covariantie-benadering. Voor vragen naar leervorderingen is de verschillscore-benadering geschikt. Er zijn echter ook onderzoekers die gebruik maken van de covariantie-benadering om leerwinst te analyseren, zie bijvoorbeeld Hoffer, Greeley en Coleman (1985), Coleman en Hoffer (1987), Brandsma (1993), Smith (1993) en Weide (1995). In *High School and Beyond*, de dataset die Coleman en zijn collega's gebruikten, blijkt dat het verschil tussen begin- en eindmeting groter is op katholieke dan op openbare scholen. Op bijvoorbeeld een leestoets scoren leerlingen op openbare en katholieke scholen bij de eerste meting respectievelijk 9,2 en 10,6. Bij de tweede meting zijn de scores 10,2 en 11,8. De gemiddelde verschillscore is dus 1,0 voor openbare scholen en 1,2 voor katholieke scholen. Voor wiskunde zijn de gemiddelde verschillscores 1,3 voor openbare en 2,3 voor katholieke scholen (zie voor de gemiddelden: Willms, 1985). Deze *achievement growth* analyseren zij vervolgens met een covariantie-analytisch model (Hoffer, Greeley & Coleman, 1985). Met als gevolg dat in de analyses van deze onderzoekers leerprestaties en -vorderingen voortdurend met elkaar worden verward.

Tegen beide methoden zijn verschillende bezwaren gemaakt. Zo staat bijvoorbeeld de betrouwbaarheid van verschillscores ter discussie (Allison, 1990; Bosker, 1991; Collins, 1996; Swint, 1994; Williams & Zimmerman, 1996). Van den Bergh en Kuhlemeier (1997) brengen naar voren dat het covariantie-analytisch model gegevens enigszins kan vertekenen. Random effecten van een covariaat beïnvloeden de grootte en de significantie van variantieschattingen (Reichardt, 1979; Snijders & Bosker, 1999).ⁱ Verder geldt voor het covariantie-analytisch model dat random hellingen niet per se wijzen op differentiële effecten tussen klassen en scholen. De variantie in hellingen zou ook het gevolg kunnen zijn van bodem- en plafond-effecten of van verschillen tussen klassen in de betrouwbaarheid van de beginmeting (Van den Bergh & Kuhlemeier, 1997).

Belangrijker dan deze technische bezwaren is echter dat er inhoudelijke verschillen zijn tussen beide methoden. Afhankelijk van de vraag moet een onderzoeker een van de methoden kiezen. Wanneer een onderzoeker geïnteresseerd is in welke leerlingen op een bepaald moment een voorsprong of achterstand hebben in het onderwijs, dan is een covariantie-analyse gewenst. Wanneer een onderzoeker wil weten welke leerlingen voorop of achterop raken in het onderwijs, dan moet er worden gekozen voor een analyse van leerwinst of relatieve posities. Uit een vergelijking van de modellen kan niet worden opgemaakt dat het ene model beter is dan het andere. Wat blijkt is dat de modellen verschillen en dat ze elkaar aanvullen.

Het verschil tussen de covariantie-analyse en de analyse van verschillen in relatieve posities, kan worden gedemonstreerd met gegevens uit VOCL'93. Met gegevens uit dat cohort wil ik nagaan of etniciteit invloed heeft op leerprestaties (covariantie-analyse) en leervorderingen (analyse van verschillen in relatieve posities). In Tabel 5.2 staan voor autochtone en allochtone leerlingen de gestandaardiseerde toetsscores tekstbegrip weergegeven.

In de psychometrie en statistiek is veel aandacht besteed aan het onderscheid tussen het analyseren van enerzijds verschilsscores en anderzijds *posttest* metingen met de *pretest* als covariaat. Dat deze twee analysemethodes in het geval van niet-experimentele designs (dus observationele of quasi-experimentele designs) tot verschillen kunnen leiden staat bekend als de paradox van Lord (1960; 1967).

Tabel 5.2: Toetsscores en verschilscore Nederlands (tekstbegrip) naar etniciteit

	AUTOCHTOON	ALLOCHTOON
Leerjaar 1	0,03	-0,63
Leerjaar 3	0,04	-0,62
Versilsscore	0,01	0,01

In beide leerjaren scoren autochtone leerlingen iets boven het gemiddelde. Allochtone leerlingen daarentegen scoren meer dan drievijfde standaarddeviatie onder het gemiddelde. De verschilscore (leerjaar 3 – leerjaar 1) is voor beide groepen vrijwel gelijk aan nul. In beide leerjaren nemen beide groepen dezelfde positie in. De achterstand in Nederlands van allochtone leerlingen op autochtone leerlingen verandert niet in de onderbouw van het voortgezet onderwijs.

Voor leervorderingen ziet de regressievergelijking met de verschilscore als afhankelijke variabele en etniciteit als onafhankelijke variabele er als volgt uit: $Y(\text{Versilsscore}) = 0,01 + 0,00 * X_1(\text{Etniciteit})$. Etniciteit heeft dus geen effect op leervorderingen. Uit Tabel 5.2 kon dit ook al worden opgemaakt, want de gemiddelde verschilscore is voor beide groepen gelijk. Etniciteit kan dus uit de regressievergelijking worden gelaten. Deze variabele draagt niets bij aan de verklaring van verschillen in relatieve posities.

De regressievergelijking voor leerprestaties met de toetsscore in leerjaar drie als afhankelijke variabele, etniciteit als onafhankelijke variabele en de toetsscore in leerjaar één als covariaat ziet er als volgt uit: $Y(\text{Toetsscore leerjaar 3}) = 0,02 - 0,30 * X_1(\text{Etniciteit}) + 0,54 * X_1(\text{Toetsscore leerjaar 1})$. Etniciteit levert, net als de eerdere toetsscore, een significante bijdrage aan de verklaring van verschillen in leerprestaties. Dit effect moet als volgt worden opgevat. Om de prestaties in leerjaar drie te voorspellen zijn prestaties van twee jaar eerder geschikt.

Leerlingen die in de brugklas goed presteerden op Nederlands, presteren twee jaar later gemiddeld genomen wederom goed. Deze voorspelling is niet perfect. Vandaar dat er variantie onverklaard blijft. Door de variabele etniciteit wordt een deel van die variantie gebonden.

Voor autochtone leerlingen ziet de regressievergelijking in de covariantie-benadering er als volgt uit: $Y(\text{Toetsscore leerjaar } 3) = 0,02 + 0,54 * X_1(\text{Toetsscore leerjaar } 1)$. Voor autochtone leerlingen is de waarde van de dummy voor etniciteit gelijk aan nul. Vandaar dat het effect van etniciteit uit de vergelijking wegvalt. Voor allochtone leerlingen is de waarde van de dummy gelijk aan één en valt de regressievergelijking om te rekenen tot: $Y(\text{Toetsscore leerjaar } 3) = -0,28 + 0,54 * X_1(\text{Toetsscore leerjaar } 1)$. Gegeven het effect van de toetsscore in leerjaar één scoren allochtone leerlingen twee jaar later 0,30 standaarddeviatie lager op de toets dan autochtone leerlingen. Beide benaderingen blijken dus duidelijk van elkaar te verschillen. De oplossing voor Lords paradox is het inzicht dat beide analysemethoden een antwoord geven op verschillende vragen, zie bijvoorbeeld Maxwell en Delaney (1990).

5.3 Het individuele niveau

In deze paragraaf ga ik na of er effecten van leerling- en gezinskenmerken zijn op de verschillen in relatieve posities voor Nederlands en wiskunde. Als eerste stap verricht ik een multiniveau analyse van de structurele kenmerken (cirkel vijf) op de verschilscores Nederlands en wiskunde. Vervolgens voeg ik steeds een concentrische cirkel toe en krijg ik uiteindelijk een model met de kenmerken van alle cirkels op de toetsscores. De kenmerken uit een cirkel worden simultaan toegevoegd. De resultaten van de multiniveau analyse zijn weergegeven in enerzijds Tabel 5.3 en 5.4 en anderzijds Bijlage 5A en 5B. In deze analyses zal ik dezelfde kenmerken opnemen als in hoofdstuk 3. De enige uitzondering is dat ik de eerdere prestaties niet als onafhankelijke variabelen gebruik. Er ontstaan namelijk interpretatieproblemen, wanneer leerwinst een functie is van eerdere prestaties (Allison, 1990).

In hoofdstuk 3 is gebleken dat in het covariantie-analytisch model talrijke leerling- en gezinskenmerken van invloed zijn op de prestaties Nederlands en wiskunde. De vraag is of deze kenmerken ook van invloed zijn op de leervorderingen op beide vakken. De verwachting is dat de resultaten van de verschillende modellen tot verschillende uitkomsten zullen leiden. Kenmerken die van invloed zijn op het niveau dat leerlingen in leerjaar drie op Nederlands en wiskunde hebben, hoeven niet

per definitie een invloed te hebben op de leerwinst die leerlingen tussen leerjaar één en drie behalen.

Structurele leerling- en gezinskenmerken

Het is de verwachting dat in de analyse van verschillen in relatieve posities *structurele kenmerken* nauwelijks of geen rol spelen (vergelijk Kuiper & Veenstra, 1995). De reden hiervoor is dat kenmerken als de sociaal-economische status, de etniciteit en het aantal ouders vooral invloed zullen hebben op de positie die leerlingen bij instroom in het voortgezet onderwijs innemen en niet op de vorderingen die ze in de onderbouw van dat onderwijs maken (Dronkers, 1997a). Dergelijke structurele kenmerken zijn vooral belangrijk in de voor- en vroegschoolse periode. De invloed van deze structurele kenmerken is dus al in de voormeting verdisconteerd, waardoor er als het ware vanzelf voor zulke kenmerken wordt gecontroleerd.ⁱⁱ

Als een structureel kenmerk een effect heeft op leervorderingen, dan is de kans het grootst dat het een sekse-effect is. Tussen jongens en meisjes zijn er verschillen in prestatiemotivatie, inzet, leefpatroon en schoolbeleving. Deze verschillen hebben mogelijk in de onderbouw van het voortgezet onderwijs hun uitwerking.

Culturele leerling- en gezinskenmerken

De *culturele kenmerken* hebben naar verwachting meer effect op leerwinst Nederlands en wiskunde dan de structurele kenmerken. Van sommige culturele kenmerken is de invloed mogelijk al verdisconteerd in het beginniveau van leerlingen. Voor de meeste kenmerken geldt echter dat ze juist tijdens het voortgezet onderwijs van belang kunnen zijn. Hierbij wordt gekeken naar de invloed van de autoritatieve opvoedingsstijl, het aantal contacturen tussen ouders en kind (cirkel vier), de huiswerkcontrole, praten met ouders (cirkel drie), het leefpatroon van leerlingen, het zelfbeeld (cirkel twee), de mening over klasgenoten, docenten en regels, de prestatiemotivatie, het aantal dagen huiswerk, het uitstellen van huiswerk en de non-verbale intelligentie (cirkel één).

De autoritatieve opvoedingsstijl, het aantal contacturen en het praten met ouders hebben naar verwachting een positief effect op de verschillen in relatieve posities op beide vakken. Voor het effect van de huiswerkcontrole is het interessant om na te gaan of dit effect ook optreedt in een analyse van verschillen in relatieve posities. In hoofdstuk 3 heb ik beweerd dat bij huiswerkcontrole vaak gesproken kan worden van crisis-interventie en dat deze vorm van interventie ongunstig zou zijn voor de prestaties van leerlingen. In de covariantie-analyse leidt het negatieve effect van huiswerkcontrole tot de vraag over wat oorzaak en wat gevolg is. Moeten ouders deze controle nalaten? Of zou het zo kunnen zijn dat ouders meer gaan controleren naarmate de schoolprestaties van hun kind lager zijn of lager worden? Wan-

neer uit de analyse van verschilscores blijkt dat huiswerkcontrole wederom een negatief effect heeft, dan wint de causale gedachte enigszins aan steun dat dit effect inhoudt dat leerlingen waarvan de ouders het huiswerk controleren meer achteruitgaan in hun relatieve positie ten opzichte van leerlingen waarvan de ouders die controle achterwege laten. Bij deze interpretatie blijft echter voorzichtigheid geboden.

De verwachting is dat de effecten van het leefpatroon van leerlingen op verschillen in relatieve posities dezelfde richting opgaan als in hoofdstuk 3. Naarmate leerlingen meer tijd op straat rondhangen, aan een bijbaantje of uitgaan besteden, wordt hun relatieve positie waarschijnlijk lager. Van een tijdbesteding waarin lezen en hobby's een belangrijke rol spelen gaat waarschijnlijk juist een positief effect uit. Dit effect is naar verwachting sterker voor Nederlands. Op de vorderingen in de onderbouw van het voortgezet onderwijs heeft het zelfbeeld waarschijnlijk ook een positief effect.

De mening over klasgenoten, docenten en regels en de prestatiemotivatie geven aan hoe leerlingen hun tijd in het voortgezet onderwijs beleven. Naarmate deze beleving positiever is, zal de relatieve positie waarschijnlijk hoger worden. Het aantal dagen huiswerk en het uitstellen van huiswerk zijn indicaties van de inzet en werkhouding van leerlingen. De verwachting is dat deze kenmerken een effect hebben op de verschillen in relatieve positie. Ook van de non-verbale intelligentie wordt een effect verwacht. Naarmate leerlingen beter kunnen abstraheren en redeneren, zullen ze waarschijnlijk meer vorderingen boeken.

Samenvattend wordt gesteld dat leerlingen meer vorderingen boeken in de onderbouw van het voortgezet onderwijs, wanneer leerlingen (1) meer autoritatief worden opgevoed; (2) meer contacturen hebben met hun ouders; (3) minder vaak door hun ouders worden aangespoord tot of gecontroleerd op het maken van huiswerk (crisis-interventie); (4) vaker met hun ouders praten (groei-interventie); (5) minder vrije tijd besteden op straat met vrienden; (6) minder vrije tijd besteden aan een bijbaantje; (7) minder vrije tijd besteden aan uitgaan; (8) een positiever zelfbeeld hebben; (9) positiever oordelen of hun medeleerlingen; (10) positiever oordelen over hun docenten en de schoolregels; (11) een hogere prestatiemotivatie hebben; (12) het maken van huiswerk minder uitstellen; (13) meer dagen in de week huiswerk maken; (14) een hogere non-verbale intelligentie hebben. Voor de vorderingen op Nederlands wordt verwacht dat (15) leerlingen meer vooruitgaan naarmate ze meer vrije tijd besteden aan lezen en hobby's. Verder wordt getoetst of er verschillen zijn in leervorderingen naar (16) sekse. Hierbij is onduidelijk welke richting deze effecten zullen hebben. Vandaar dat tweezijdig wordt getoetst. Van de overige structurele kenmerken wordt geen effect verwacht op leer-

vorderingen, te weten de sociaal-economische status, de etniciteit, de gezinssamenstelling en de positie in de kinderrij.

Resultaten

In de analyse van verschillen in relatieve posities heb ik de structuur van concentrische cirkels uit hoofdstuk 3 gehanteerd, zie dat hoofdstuk voor een beschrijving van de data. Net als in de voorgaande hoofdstukken zijn alle variabelen op ordinaal of hoger meetniveau gestandaardiseerd. Ook de verschillscores uit Tabel 5.1 zijn omgezet in z-scores. Hierdoor zijn de regressiecoëfficiënten gelijk aan bèta-gewichten. De nominale variabelen, zoals de sekse of de positie in de kinderrij, zijn niet gestandaardiseerd. Hierdoor is, net als in de vorige hoofdstukken, de totale variantie niet precies gelijk aan één.

Allereerst heb ik gekeken naar de effecten van structurele kenmerken (cirkel vijf) op de verschillen in relatieve posities. In Bijlage 5A en 5B heb ik de resultaten voor Nederlands en wiskunde apart weergegeven. Onder constant houding van de andere kenmerken, hebben jongens op Nederlands een verschillscore die 0,194 lager is dan die van meisjes. Op wiskunde is opvallend genoeg de verschillscore van jongens 0,057 lager dan die van meisjes. Op beide vakken krijgen meisjes dus een hogere relatieve positie dan jongens, waarbij het effect voor Nederlands veel sterker is. Voor de rest zijn er geen invloeden op de verschillen in relatieve positie. Het enige structurele kenmerk met een effect op verschillen in relatieve positie is de sekse.

Aan het model met structurele kenmerken heb ik vervolgens cirkel voor cirkel culturele kenmerken toegevoegd. Van de kenmerken uit cirkel vier heeft de autoritatieve opvoedingsstijl geen effect op de verschillen in relatieve positie op beide vakken. Het aantal contacturen met ouders heeft een positief effect op zowel Nederlands als wiskunde. De twee kenmerken uit cirkel drie hebben een invloed op beide vakken. Leerlingen die vaker door hun ouders worden aangespoord tot of gecontroleerd op het maken van huiswerk boeken minder leervorderingen. Voor praten met ouders geldt dat het effect in beide gevallen positief is. De meeste algemene leerlingkenmerken (cirkel twee) blijken een effect te hebben op de verschillen in relatieve posities. Leerlingen boeken op beide vakken meer vooruitgang, wanneer ze een positief zelfbeeld hebben. Ze gaan meer achteruit op Nederlands en wiskunde als ze veel op straat rondhangen. Van een bijbaantje blijkt alleen op Nederlands een negatief effect uit te gaan. Voor uitgaan is juist alleen van een effect op wiskunde sprake. Voor deze twee kenmerken van het leefpatroon zijn de effecten op het andere vak niet significant, maar wel in dezelfde richting. Door de opname van de algemene leerlingkenmerken is de

invloed van het aantal contacturen niet langer significant op Nederlands. Op wiskunde is het effect van dit kenmerk gehalveerd.

Tot slot zijn de schoolse leerlingkenmerken (cirkel één) in de analyse opgenomen. Van deze kenmerken hebben de mening over docenten en regels en de prestatiemotivatie voor beide vakken een positieve invloed op de verschillen in relatieve posities. Verder geldt dat naarmate leerlingen meer dagen in de week huiswerk maken, hun vooruitgang in wiskunde groter is. De overige kenmerken uit deze cirkel hebben geen invloed. Door de opname van cirkel één is het effect van het zelfbeeld meer dan gehalveerd. Het effect van het zelfbeeld op de leervorderingen wordt gemedieerd door de mening over docenten en regels en de prestatiemotivatie. Ook de effecten van de andere algemene leerlingkenmerken zijn minder sterk geworden.

*Tabel 5.3: Multiniveau analyse op leervorderingen. Spaarzaam model met contrasten (chi-kwadraat verdeeld, $df=1$) met **verschillende** of **vergelijkbare** effecten*

	CONTRAST	O ²	NED	WIS
Sekse (1=jongen)	-0,15	12,61	-0,168 (0,026) ***	-0,021 (0,027)
Huiswerkcontrole	-0,02	2,17	-0,030 (0,009) ***	-0,030 (0,009) ***
Op straat met vrienden	0,01	0,16	-0,041 (0,010) ***	-0,041 (0,010) ***
Bijbaan	-0,03	2,77	-0,028 (0,009) **	-0,028 (0,009) **
Uitgaan	0,02	1,52	-0,023 (0,009) **	-0,023 (0,009) **
Zelfbeeld	0,00	0,06	0,029 (0,010) **	0,029 (0,010) **
Mening over docenten/regels	0,01	0,47	0,046 (0,010) ***	0,046 (0,010) ***
Prestatiemotivatie	0,02	0,65	0,039 (0,010) ***	0,039 (0,010) ***
Aantal dagen huiswerk	-0,05	8,28	0,019 (0,013)	0,067 (0,014) **
Constante			0,100 (0,031)	-0,017 (0,039)

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

In Tabel 5.3 geef ik een spaarzaam model weer. Bij het opstellen van dit model heb ik de niet-significante variabelen uit de analyse verwijderd. Vervolgens ben ik voor de overgebleven variabelen nagegaan of ze een *vergelijkbaar of verschillend effect* hebben op leervorderingen Nederlands en wiskunde. In Tabel 5.3 geef ik het verschil tussen elk paar van effecten weer. Het contrast is berekend door van de parameter voor Nederlands de parameter voor wiskunde af te trekken. Het contrast is dus groter dan nul als het effect op de verschillscore Nederlands groter is dan het effect op de verschillscore wiskunde. Omgekeerd is het contrast kleiner

dan nul als het effect op wiskunde groter is dan het effect op Nederlands. Het verschil tussen deze parameters kan worden getoetst met een chi-kwadraattoets. De kritieke waarde van chi-kwadraat bij een significantieniveau van 0,05 en één vrijheidsgraad is 3,84.

Tabel 5.4: Multiniveau analyse op leervorderingen. Variantie- en covariantiecomponenten en modelfit

	VARIANTIE NEDERLANDS	VARIANTIE WISKUNDE	COVARIANTIE	DEVIANCE)
LEEG MODEL	PAR. (S.E.)	PAR. (S.E.)	PAR. (S.E.)		
Leerlingniveau	0,879 (0,016)	0,843 (0,016)	0,109 (0,012)	34.278,0	
Klasniveau	0,067 (0,011)	0,067 (0,011)	0,019 (0,008)		
Schoolniveau	0,058 (0,013)	0,130 (0,022)	0,020 (0,012)		
CIRKEL 5					
Leerlingniveau	0,870 (0,016)	0,839 (0,016)	0,105 (0,012)	34.179,8	98,2
Klasniveau	0,063 (0,010)	0,067 (0,011)	0,018 (0,008)		(df=20)
Schoolniveau	0,063 (0,013)	0,125 (0,021)	0,022 (0,012)		p<0,001
CIRKEL 4					
Leerlingniveau	0,869 (0,016)	0,836 (0,016)	0,104 (0,012)	34.156,6	23,2
Klasniveau	0,062 (0,010)	0,068 (0,011)	0,018 (0,008)		(df=4)
Schoolniveau	0,063 (0,013)	0,125 (0,021)	0,021 (0,012)		p<0,001
CIRKEL 3					
Leerlingniveau	0,865 (0,016)	0,834 (0,016)	0,101 (0,012)	34.118,7	37,9
Klasniveau	0,062 (0,010)	0,068 (0,011)	0,018 (0,008)		(df=4)
Schoolniveau	0,063 (0,013)	0,125 (0,021)	0,022 (0,012)		p<0,001
CIRKEL 2					
Leerlingniveau	0,859 (0,016)	0,828 (0,016)	0,095 (0,012)	34.028,4	90,3
Klasniveau	0,061 (0,010)	0,069 (0,011)	0,018 (0,008)		(df=10)
Schoolniveau	0,065 (0,013)	0,127 (0,022)	0,024 (0,012)		p<0,001
CIRKEL 1					
Leerlingniveau	0,852 (0,015)	0,822 (0,016)	0,089 (0,012)	33.939,6	88,8
Klasniveau	0,060 (0,010)	0,065 (0,011)	0,017 (0,008)		(df=12)
Schoolniveau	0,066 (0,013)	0,128 (0,022)	0,026 (0,012)		p<0,001
SPAARZAAM MODEL					
Leerlingniveau	0,855 (0,015)	0,826 (0,016)	0,091 (0,012)	33.987,6	48,0
Klasniveau	0,059 (0,010)	0,064 (0,011)	0,016 (0,008)		(df=39)
Schoolniveau	0,066 (0,013)	0,131 (0,022)	0,025 (0,013)		p>0,100

Twee effecten verschillen naar vak, namelijk sekse en aantal dagen huiswerk. Deze effecten zijn vet weergegeven. De andere variabelen hebben een vergelijkbaar effect op beide toetsen. Voor deze effecten kan daarom één parameter worden geschat. Hierbij gaat het om de volgende effecten: huiswerkcontrole (cirkel drie), op straat met vrienden, bijbaan, uitgaan, zelfbeeld (cirkel twee), mening over docenten en regels en prestatiemotivatie (cirkel één). Van de tien structurele kenmerken heeft alleen het verschil tussen jongens en meisjes invloed. Van de veertien culturele kenmerken hebben acht een invloed.

Variantiecomponenten en modelfit

In Tabel 5.4 vergelijk ik de variantiecomponenten en de modelfit. In het lege model bedragen de variantiecomponenten voor verschillen in relatieve positie Nederlands op leerlingniveau 0,879, op klasniveau 0,067 en op schoolniveau 0,058. De intra-klasse correlatiecoëfficiënt voor verschillen tussen klassen en scholen in verschillen in relatieve positie Nederlands is $(0,067 + 0,058) / (0,879 + 0,067 + 0,058) = 0,13$. De variantie tussen klassen en scholen bedraagt dus 13 procent. Deze variantie betreft voor 54 procent variantie tussen klassen en voor 46 procent variantie tussen scholen. Voor verschillen tussen scholen ziet de intra-klasse correlatiecoëfficiënt er als volgt uit: $0,058 / (0,067 + 0,058) = 0,46$. Voor verschillen in relatieve positie wiskunde is de variantie tussen klassen en scholen 19 procent en tussen scholen 66 procent. Met de schattingen van de varianties en covarianties kan op elk niveau de correlatie tussen de verschillen in relatieve positie op beide vakken worden berekend. In het lege model is de correlatie 0,23 op school-, 0,28 op klas- en 0,13 op leerlingniveau. De samenhang tussen de verschillen in relatieve positie voor Nederlands en wiskunde is dus op alle niveaus zwak.

In het tweede model zijn de structurele kenmerken toegevoegd. Op leerlingniveau verklaren deze kenmerken ongeveer 1 procent van de variantie in verschillen in relatieve positie op beide vakken. Op klas- en schoolniveau verklaren de structurele kenmerken voor wiskunde 2,2 en 3,0 procent van de variantie. Voor Nederlands wordt er op deze niveaus niets verklaard. Doordat de variantie op schoolniveau toeneemt, is de verklaarde variantie zelfs negatief (de formules 2.13, 2.14 en 2.15 zijn gebruikt; zie verder: Snijders & Bosker, 1999, hoofdstuk 7). In het model met alle culturele kenmerken, dus tot en met cirkel één, is de verklaarde variantie op leerlingniveau toegenomen tot ongeveer 2,5 procent. Voor wiskunde is de verklaarde variantie op de andere niveaus rond de 2 procent. Voor Nederlands is de verklaarde variantie op klasniveau 0,7 en die op schoolniveau is nog steeds negatief. Slechts een zeer beperkt deel van de variantie in verschillen in relatieve positie blijkt te binden door effecten van leerlingen en gezinnen.

In hoofdstuk 3 heb ik de verklaarde variantie grafisch weergegeven (Figuur 3.1 en 3.2). Voor de verschillscore-benadering zien die figuren er totaal anders uit. De variantie op individueel niveau loopt voor beide vakken op van iets minder dan 1 procent (cirkel vijf) naar ongeveer 2,5 procent (cirkel één). Op klasniveau schommelt de variantie rond de 0 en de 2 procent voor respectievelijk Nederlands en wiskunde. Op schoolniveau is de variantie voor Nederlands negatief. Voor wiskunde neemt de variantie op dat niveau af van 3,0 (cirkel vijf) naar 1,8 procent (cirkel één).

In Tabel 5.4 valt te zien dat de onverklaarde variantie op het individuele niveau groot is. Met de leerling- en gezinskenmerken valt deze onverklaarde variantie moeilijk lager te krijgen. Wanneer in de verschillscore-benadering wordt gekeken naar betrouwbaarheid zoals gemeten door bijvoorbeeld Cronbachs alfa, dan blijkt dat deze coëfficiënt betrekkelijk laag is. Op het individuele niveau is er een grote restvariantie die betekenisloze toevalvariatie is. Deze ruis is een deel van de variantie die niet kan en hoeft te worden verklaard. Eigenlijk zou de verklaarde variantie als fractie van de betrouwbare variantie moeten worden bepaald. Stel dat de betrouwbaarheid van de verschillscore \forall is, dan is een fractie van de variantie van $1 - \forall$ onbetrouwbaar. Wanneer de betekenisvolle variantie \forall is, dan kan de verklaarde variantie nooit meer dan \forall zijn en kan de multiniveau analyse beter worden beoordeeld door het percentage verklaarde variantie gedeeld door de betrouwbare variantie: R^2 / \forall .

Daarnaast is het de vraag hoe relevant de variabiliteit in verschillscores is. Om de relevantie van de verschillscores te bepalen, kan de voor onbetrouwbaarheid gecorrigeerde correlatie tussen begin- en eindmeting worden berekend. Voor de verschillscore in wiskunde wil ik dit illustreren. De correlatie tussen de toetsen voor wiskunde is 0,74. Om het voorbeeld eenvoudig te houden wordt de betrouwbaarheid van beide toetsen op 0,84 gesteld. De scores op de beginmeting (Y_1) en eindmeting (Y_2) zijn op te delen in de ware scores (T) en de random error (E). De bijbehorende variantie is: $\sigma_Y^2 = \sigma_T^2 + \sigma_E^2$. De variantie in de begin- of eindmeting wordt bepaald door de ware en de foutenvariantie. Deze foutenvariantie is gelijk aan $1 - 0,84 = 0,16$. Aangezien de begin- en eindmeting gestandaardiseerd zijn, is de covariantie tussen beide toetsen gelijk aan de correlatie. Tevens valt af te leiden dat de ware covariantie gelijk is aan de geobserveerde covariantie en dat de ware variantie 0,84 is, het verschil tussen de variantie van de toetsscores (1,00) en de foutenvariantie (0,16). Met deze waarden kan de ware correlatie tussen de begin- en eindmeting worden berekend: $0,74 / \sqrt{(0,84 \cdot 0,84)} = 0,88$. Verder geldt dat de variantie van de ware verschillscore 0,20 is (variantie $(T_1 - T_2) =$ variantie $T_1 +$ variantie $T_2 - 2 \cdot$ covariantie $(T_1, T_2) = 0,84 + 0,84 - 2 \cdot 0,74$). De standaarddeviatie van de ware verschillscore is dan 0,45. Ter vergelijking de standaarddeviatie van de

ware begin- en eindmeting is $\sqrt{0,84} = 0,92$. Uit deze uiteenzetting blijkt dat de ware standaarddeviatie in de verschillscores twee keer zo klein is als de ware standaarddeviatie in de begin- of eindmeting. Toch is een standaarddeviatie van 0,45 nog vrij groot. Het betekent immers dan een leerling die in het eerste leerjaar op het gemiddelde scoort, twee jaar later heel goed 0,45 boven of onder het gemiddelde kan scoren. Wanneer een leerling twee standaarddeviaties boven of het onder gemiddelde scoort is het verschil zelfs 0,90. Een leerling die in de brugklas op mavo-niveau scoort en die twee jaar later 0,45 of zelfs 0,90 boven het gemiddelde scoort, presteert in leerjaar drie dan ongeveer op havo- respectievelijk vwo-niveau. Dergelijke verschillen tussen leerjaar één en drie zijn van groot gewicht.ⁱⁱⁱ

Door de opname van extra cirkels blijkt de deviance wel steeds af te nemen. De afname in deviance is in alle gevallen significant. Ofschoon de verklaarde variantie laag is, leidt de opname van een extra cirkel dus steeds tot een verbetering van de verklaring van de verschillen in relatieve positie voor beide vakken. Verder blijkt dat leerlingkenmerken meer invloed hebben op leervorderingen dan gezinskenmerken. Eerder werd al geconstateerd dat structurele kenmerken (vrijwel) geen en culturele kenmerken een zwakke invloed hebben op de vooruitgang op beide vakken.

Na controle voor leerlingkenmerken zijn de correlaties tussen de verschillen in relatieve positie voor beide vakken enigszins veranderd. In het spaarzame model bedragen deze op leerling-, klas- en schoolniveau 0,11, 0,27 en 0,26. Deze correlaties zijn laag vergeleken met die in de covariantie-benadering. Verder blijken voor beide vakken de residuele intra-klasse correlatiecoëfficiënten vrijwel gelijk aan de coëfficiënten in het lege model.

5.4 Het klas- en schoolniveau

Voor verschillen in leerprestaties bieden leerling- en gezinskenmerken een goede verklaring. Voor verschillen in leervorderingen geldt dat veel minder. De vraag is nu wat de invloed op leervorderingen is van kenmerken op klas- en schoolniveau. Net als in hoofdstuk 4 probeer ik deze vraag op verschillende manieren te beantwoorden. Allereerst ga ik na wat het effect is van de toevoeging van het door leerlingen gevolgde *schooltype*. Vervolgens onderzoek ik of er *random effecten* zijn van leerling- en gezinskenmerken. In hoofdstuk 4 bleek een aantal kenmerken een differentieel effect te hebben. Mogelijk dat er voor leervorderingen ook dergelijke

effecten worden gevonden. Tot slot voeg ik *vaste effecten* van contextuele en schoolkenmerken toe (zie voor een beschrijving van deze gegevens: hoofdstuk 4).

Controles op klas- en schoolniveau

Aan de kenmerken van leerlingen en gezinnen heb ik net als in hoofdstuk 4 het *schooltype* van leerlingen toegevoegd. Het is op voorhand onduidelijk welke kant deze effecten opgaan.

*Tabel 5.5: Multiniveau analyse op leervorderingen.
Model met toevoeging van schooltype*

	NEDERLANDS	WISKUNDE
Sekse (1=jongen)	-0,159 (0,026) ***	-0,019 (0,027)
Huiswerkcontrole	-0,025 (0,009) **	-0,025 (0,009) **
Op straat met vrienden	-0,034 (0,010) ***	-0,034 (0,010) ***
Bijbaan	-0,024 (0,009) **	-0,024 (0,009) **
Uitgaan	-0,021 (0,009) *	-0,021 (0,009) *
Zelfbeeld	0,022 (0,010) *	0,022 (0,010) *
Mening over docenten/regels	0,045 (0,010) ***	0,045 (0,010) ***
Prestatiemotivatie	0,045 (0,010) ***	0,045 (0,010) ***
Aantal dagen huiswerk	0,006 (0,014)	0,048 (0,014) ***
Schooltype (Referentiegroep: mavo)		
- ivbo	0,036 (0,068) } ***	-0,498 (0,073) } ***
- vbo	-0,135 (0,043) } ***	-0,115 (0,038) } ***
- vbo/mavo	0,095 (0,105) } ***	0,331 (0,114) } ***
- havo	0,172 (0,046) } ***	-0,048 (0,051) } ***
- havo/vwo	0,307 (0,108) } ***	-0,056 (0,114) } ***
- vwo	0,263 (0,049) } ***	0,007 (0,054) } ***
Constante	0,075 (0,040)	0,063 (0,046)

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

De resultaten zijn in Tabel 5.5 weergegeven. Vergeleken met mavo-leerlingen hebben leerlingen in de hogere schooltypen gemiddeld genomen een verschilscore voor Nederlands die 0,172 tot 0,307 hoger is. Van de leerlingen in de lagere schooltypen lopen vbo-leerlingen relatief gezien een achterstand op beide vakken op bij mavo-leerlingen. De verschilscore van vbo-leerlingen is voor Nederlands 0,135 en voor wiskunde 0,115 lager. Ivbo-leerlingen boeken beduidend minder

voortgang op wiskunde dan mavo-leerlingen. Het verschil is een halve standaarddeviatie. Vbo/mavo-leerlingen daarentegen boeken gemiddeld genomen meer vorderingen voor wiskunde dan mavo-leerlingen. Naar schooltype zijn er dus verschillen in gemiddelde verschilscore. Deze verschillen volgen niet een heel duidelijk patroon.

Uit de simultane toets (Goldstein, 1995²) voor de zes dummy's voor schooltype blijkt dat het contrast 75,11 is met zes vrijheidsgraden, wat overeenkomt met een significantieniveau kleiner dan 0,001. Hieruit kan worden opgemaakt dat het schooltype dus een verschillend effect heeft op de leervorderingen op beide vakken. In Tabel 5.5 zijn de effecten die verschillend zijn voor beide vakken met vet aangegeven.

Door de opname van het schooltype in de analyse veranderen de variantiecomponenten. De deviance neemt af met 122,8 bij twaalf vrijheidsgraden, wat significant is bij een alfa van 0,001. Deze gegevens worden in Tabel 5.8 samengevat onder het kopje 'schooltype'. In het vorige hoofdstuk werd in Figuur 4.3 en 4.4 de netto variantie in de prestaties Nederlands en wiskunde weergegeven. Voor de vorderingen op beide vakken zien de figuren er ongeveer hetzelfde uit. De relatieve binnenschoolse variantie bedraagt 86 procent voor Nederlands en 82 procent voor wiskunde. De relatieve tussenschoolse variantie voor Nederlands is 14 procent, welke valt te splitsen in een tweevijfde deel tussen klassen en een drievijfde deel tussen scholen. Voor wiskunde is de relatieve tussenschoolse variantie 18 procent, met eenderde deel op klasniveau en tweederde op schoolniveau.

In de tweede plaats ben ik nagegaan of klassen en scholen voor alle leerlingen even goed zijn. Ik heb dus gekeken naar *random effecten*. Voor alle leerling- en gezinskenmerken uit paragraaf 3 ben ik op klas- en schoolniveau nagegaan of er sprake is van enerzijds random hellingen en anderzijds covarianties tussen het random intercept en de random hellingen. Intercept-helling-covarianties heb ik niet gevonden, maar er zijn wel random hellingen.

In de analyse van verschillen in relatieve posities zijn er random hellingen voor sekse en het aantal dagen huiswerk (beide voor wiskunde), zie Tabel 5.6. Hoe belangrijk is een random helling van 0,022 voor het aantal dagen huiswerk op de verschilscore wiskunde? De standaarddeviatie van de helling is $\sqrt{0,022} = 0,148$ en de gemiddelde helling is 0,048 (zie Tabel 5.5). Wanneer er twee standaarddeviaties bij dit gemiddelde worden opgeteld of er van afgetrokken, dan is het bereik -0,248 tot 0,344. Het effect van het aantal dagen huiswerk op verschillen in relatieve positie op beide vakken is dus in sommige klassen negatief

en in andere klassen positief. Dit effect varieert ongeveer drievijfde standaarddeviatie.

*Tabel 5.6: Multiniveau analyse op leervorderingen. Random effecten**

SCHOOLNIVEAU	PAR. (S.E.)
Variantie Nederlands	0,083 (0,015)
Variantie wiskunde	0,114 (0,019)
Covariantie Nederlands-wiskunde	0,037 (0,013)
KLASNIVEAU	
Variantie Nederlands	0,052 (0,009)
Variantie wiskunde	0,041 (0,010)
Covariantie Nederlands-wiskunde	0,021 (0,007)
Random helling van sekse (wiskunde)	0,043 (0,015)
Random helling van aantal dagen huiswerk (wiskunde)	0,022 (0,006)
LEERLINGNIVEAU	
Variantie Nederlands	0,848 (0,015)
Variantie wiskunde	0,795 (0,016)
Covariantie Nederlands-wiskunde	0,085 (0,012)

* Rekening houdend met de vaste effecten uit Tabel 5.5

Voor het random effect tussen jongens en meisjes geldt dat voor wiskunde het bereik loopt van $-0,429$ tot $0,396$. Voor de wiskundevorderingen maakt het dus nogal wat uit in welke klas meisjes en jongens zitten. In sommige klassen gaan jongens meer vooruit, maar in andere klassen boeken meisjes de meeste voorsprong.

Door de opname van random effecten neemt de variantie in de verschillcores wiskunde af, zie Tabel 5.8. Bij twee vrijheidsgraden is de afname in de deviance met 29,2 significant op een niveau van 0,001. Door de opname van de random effecten veranderen de vaste effecten nauwelijks. Deze effecten worden daarom niet opnieuw weergegeven.^{iv}

Effectieve scholen

Ten slotte onderzoek ik of er vaste effecten van contextuele en schoolkenmerken zijn op de verschillen in relatieve posities voor Nederlands en wiskunde. Als eerste stap verricht ik een multiniveau analyse van de contextuele kenmerken (cirkel C) op de verschillcores op beide vakken. Vervolgens voeg ik schoolkenmerken (cirkel B) toe. De docentkenmerken (cirkel A) laat ik opnieuw buiten be-

schouwing, omdat bij de docentgegevens de nodige twijfels zijn te plaatsen, zie paragraaf 4.3. De resultaten van de multiniveau analyses zijn weergegeven in Tabel 5.7.

In hoofdstuk 4 is gebleken dat in het covariantie-analytisch model enkele contextuele en schoolkenmerken van invloed zijn op de prestaties Nederlands en wiskunde. De vraag is of deze kenmerken ook van invloed zijn op de leervorderingen op beide vakken.

Contextuele kenmerken

Het is onduidelijk of kenmerken die vooral betrekking hebben op de *compositie* van scholen een effect hebben in de verschillscore-benadering. In hoofdstuk 4 heb ik een aantal kenmerken naar voren gehaald die te maken hebben met de compositie van scholen, te weten de locatiegrootte, het percentage allochtone leerlingen, de denominatie en de toelatingscriteria. Van deze vier kenmerken blijken de eerste drie gerelateerd te zijn aan de prestaties van leerlingen. Het is maar de vraag of deze kenmerken van invloed zijn op leervorderingen. Eerder in dit hoofdstuk bleken bijvoorbeeld structurele leerling- en gezinskenmerken niet van invloed op vorderingen van leerlingen. De verklaring daarvoor was dat structurele kenmerken een effect hebben op de positie die leerlingen bij aanvang in het voortgezet onderwijs innemen. De invloed van dat soort kenmerken is al in de voormeting verdisconteerd. In de verschillscore-benadering wordt er dus eigenlijk automatisch rekening mee gehouden. Voor de contextuele kenmerken zou hetzelfde kunnen gelden. Op scholen met kleine locaties, op ‘witte’ scholen of op bijvoorbeeld katholieke scholen zitten mogelijk ander soort leerlingen dan op scholen met veel leerlingen op één locatie, op scholen met veel allochtone leerlingen of op bijvoorbeeld openbare scholen. Dit soort verschillen uit zich, zo zou kunnen worden verondersteld, vooral in het aanvangsniveau van leerlingen. De vorderingen die leerlingen maken staan er mogelijk los van. Als dit scenario empirisch wordt bevestigd, dan hoeven ouders zich bij de schoolkeuze niet druk te maken om de contextuele verschillen tussen scholen. Er zou dan gelden dat leerlingen evenveel leren op scholen met een gunstige als op scholen met een ongunstige compositie.

Het zou echter ook kunnen dat de locatiegrootte, het percentage allochtone leerlingen en de denominatie wel een effect hebben op de vorderingen van leerlingen. De verklaring daarvoor zou zijn dat dit soort kenmerken niet alleen van invloed zijn op het aanvangsniveau van leerlingen, maar ook op processen binnen de school. Een negatief effect van de locatiegrootte op vorderingen zou het gevolg kunnen zijn van de grootschaligheid op veel scholen voor voortgezet onderwijs. Leerlingen boeken mogelijk minder vooruitgang op scholen waar ze een ‘nummer’ zijn en al snel verloren gaan in de massa. Een negatief effect van het percen-

tage allochtone leerlingen zou veroorzaakt kunnen worden doordat docenten hun lessen aanpassen of door vergelijkings- en beïnvloedingsprocessen tussen leerlingen. Voor de denominatie gelden soortgelijke vragen. Zijn verschillen tussen denominaties toe te schrijven aan selectiemechanismen (Witte, 1992) en is er dus enkel een relatie met leerprestaties? Of zijn er ook denominatie-effecten op leervorderingen, bijvoorbeeld doordat het klimaat op dergelijke scholen geschikter is voor het leren (Coleman, Hoffer & Kilgore, 1982)?

Schoolkenmerken

Voor kenmerken die te maken hebben met *processen* in het voortgezet onderwijs geldt mogelijk dat de invloed op leerprestaties vrijwel gelijk is aan de invloed op leervorderingen. In het vorige hoofdstuk zijn er effecten gevonden van de teamstabiliteit, maatregelen bij lesuitval, het mentoraatsysteem, de docent-leerling-ratio en de aard van de brugperiode. Deze kenmerken hebben betrekking op wat er in de klas of de school gebeurt. Het zijn kenmerken waarop scholen beleid kunnen voeren en waaruit de toegevoegde waarde van scholen kan blijken. De invloed van deze schoolkenmerken is niet in de voormeting verdisconteerd, zoals dat mogelijk wel geldt voor contextuele kenmerken.

Samenvattend en in lijn met het vorige hoofdstuk, is de verwachting dan ook dat in de onderbouw van het voortgezet onderwijs door leerlingen meer vorderingen worden geboekt, wanneer: (1) de teamstabiliteit groter is; (2) lesuitval opgevangen wordt door een invalregeling of een verplichte huiswerkkles; (3) het mentoraatsysteem minder intensief is; (4) de docent-leerling-ratio kleiner is. Ook wordt getoetst of er verschillen zijn tussen scholen met (5) verschillende vormen van brugperiode. Hierbij is het niet duidelijk in welke richting dit effect zal optreden.

Deze effecten hoeven niet voor alle schooltypen hetzelfde te zijn. Net als in hoofdstuk 4 zal na worden gegaan of effecten gelden voor alle schooltypen of juist voor een deel van de schooltypen, bijvoorbeeld alleen voor scholen voor beroepsonderwijs of alleen voor scholen voor algemeen vormend onderwijs.

Overeenkomstig het vorige hoofdstuk wordt geen effect verwacht van de nadruk op basisvaardigheden, het professioneel en onderwijskundig leiderschap, de bewaking van leervorderingen, en de samenwerking tussen school en thuis. Voor dergelijke kenmerken zijn er geen aanwijzingen dat die een effect hebben op leerprestaties of -vorderingen in de onderbouw van het Nederlandse voortgezet onderwijs. Ook van het ordelijk schoolklimaat, de eenheid in visie en doelen of het prestatieniveau als extra toelatingscriterium wordt nu geen effect verwacht.

*Tabel 5.7: Multiniveau analyse op leervorderingen.
Model met toevoeging van schoolkenmerken*

	CONTRAST	O ²	NEDERLANDS	WISKUNDE
CONTEXTUELE KENMERKEN				
Locatiegrootte (log ₁₀)	-0,02	0,28	-0,070 (0,029) *	-0,070 (0,029) *
Denominatie (Referentiegroep: openbaar)				
- algemeen bijzonder	0,13	0,80	0,047 (0,098) *	0,047 (0,098) *
- rooms-katholiek	0,07	0,63	-0,050 (0,065) *	-0,050 (0,065) *
- interconfessioneel	0,28	3,67	-0,028 (0,109) *	-0,290 (0,136) *
- protestants-christelijk	0,08	0,58	0,065 (0,070) *	0,065 (0,070) *
- reformatorisch	-0,40	3,66	-0,265 (0,155) *	0,177 (0,206) *
- gereformeerd-vrijgemaakt	-0,30	2,31	0,113 (0,141) *	0,113 (0,141) *
- mavo * geref.-vrijgemaakt			0,190 (0,118) *	
SCHOOLKENMERKEN				
Maatregelen bij lesuitval (Referentiegroep: lesuitval)				
- invalregeling	0,24	5,70	0,263 (0,075) ***	0,099 (0,085) *
- verplichte huiswerkles	0,19	1,87	0,140 (0,095) ***	0,140 (0,095) *
- havo/vwo * invalregeling			-0,188 (0,084) ***	
Aard brugperiode (Referentiegroep: dakpansgewijs)				
- homogeen	-0,17	2,41	0,019 (0,071) *	0,019 (0,071) *
- heterogeen	-0,24	2,51	-0,202 (0,105) *	-0,202 (0,105) *
- vwo * heterogeen			0,420 (0,146) *	
Mentoraatsysteem (Referentiegroep: smal/verticaal)				
- smal/horizontaal	0,01	0,01	-0,061 (0,059) *	-0,061 (0,059) *
- breed/verticaal	0,10	1,12	-0,111 (0,064) *	-0,111 (0,064) *
- breed/horizontaal	-0,13	0,94	-0,262 (0,090) *	-0,262 (0,090) *
Docent-leerling ratio	0,00	0,01	0,026 (0,026) *	0,026 (0,026)
(I)vbo * docent-leerling ratio			-0,128 (0,040) *	

NOOT: Op individueel niveau zijn de kenmerken uit Tabel 5.5 meegenomen.

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Resultaten

Van de locatiegrootte gaat een invloed uit op de leerwinst op beide vakken, zie Tabel 5.7. Dit duidt erop dat de locatiegrootte van invloed is op processen binnen de school. De leervorderingen zijn lager naarmate locaties groter zijn. Van het percentage allochtone leerlingen gaat geen effect uit op leervorderingen. Verschillen tussen 'witte' en 'zwarte' scholen hangen dus samen met verschillen in in-

stroom en niet met processen binnen scholen. Dit resultaat lijkt erop te wijzen dat leerlingen op scholen met veel allochtone leerlingen alleen slechter presteren, omdat ze gemiddeld genomen een achterstand hadden toen ze aan het voortgezet onderwijs begonnen. Dit geldt voor zowel de autochtone als de allochtone leerlingen op dat soort scholen.

Een effect van de denominatie is wel aanwezig. Wanneer de negen denominatie-effecten, zie Tabel 5.7, worden weggelaten, dan neemt de deviance af met 19,7 bij acht vrijheidsgraden. Het model verslechtert hierdoor significant. Voor dit kenmerk geldt dat het niet alleen samenhangt met leerprestaties (hoofdstuk 4), maar dat er ook een invloed van uitgaat op leervorderingen. Hieruit kan worden opgemaakt dat verschillen tussen scholen van verschillende denominaties samenhangen met de processen binnen scholen. Verschillen tussen denominaties komen niet door verschillen in leerlingencompositie. Door de toevoeging van contextuele kenmerken neemt de deviance af met 41,7 bij achttien vrijheidsgraden, zie Tabel 5.8.

Voor de overige schoolkenmerken is er ook een aantal effecten gevonden, zie Tabel 5.7. Deze effecten gaan steeds in dezelfde richting en zijn vaak ook van een vergelijkbare grootte als in het covariantie-analytisch model. Ivbo-, vbo- en mavo-leerlingen boeken meer vooruitgang in Nederlands, wanneer scholen lesuitval opvangen door een invalregeling. Voor havo- en vwo-leerlingen ontbreekt dit effect.^v Scholen die lesuitval opvangen door een verplichte huiswerkles boeken gemiddeld genomen meer vooruitgang in Nederlands en wiskunde, maar dit effect is niet significant. De aard van de brugperiode resulteert ook in verschillen in vooruitgang Nederlands. Voor dat vak geldt dat leerlingen, met uitzondering van vwo-leerlingen, op scholen met een heterogene brugperiode een achterstand oplopen ten opzichte van leerlingen op scholen met een dakpansgewijze brugperiode. De effecten van de aard van de brugperiode op vorderingen in Nederlands zijn gezamenlijk significant bij een niveau van 0,05. Door verwijdering van deze kenmerken neemt de deviance namelijk af met 11,0. Op scholen met een breed-horizontaal mentoraatsysteem presteren leerlingen niet alleen slechter in leerjaar drie (hoofdstuk 4), maar boeken ze ook minder vorderingen tussen leerjaar één en drie. In het algemeen geldt, hoe uitgebreider het mentoraatsysteem, hoe minder vooruitgang leerlingen boeken. De drie effecten van het mentoraatsysteem blijken gezamenlijk significant bij een niveau van 0,05. In het vorige hoofdstuk is al de vraag naar causaliteit naar voren gekomen. Nu het effect van het mentoraatsysteem niet alleen in de covariantie- maar ook in de verschillscore-benadering negatief is, wint de causale gedachte, hoe uitgebreider het mentoraatsysteem hoe slechter de prestaties van leerlingen, aan steun.

Van de docent-leerling-ratio gaat geen effect uit op de leervorderingen van leerlingen in het algemeen vormend onderwijs. Net als in het vorige hoofdstuk geldt echter dat leerlingen in het beroepsonderwijs wel gebaat zijn bij een lagere docent-leerling-ratio. Door de toevoeging van schoolkenmerken neemt de deviance af met 66,2 bij 27 vrijheidsgraden, zie Tabel 5.8. De teamstabiliteit, een ordelijk schoolklimaat, de eenheid in visie en doelen, en prestaties als extra toelatingscriterium hebben geen effect op de vooruitgang in beide vakken.^{vi}

In Tabel 5.7 zijn alle effecten op de leervorderingen weergegeven. In de tabel is ook aangegeven of effecten voor beide vakken gelijk of verschillend zijn. Uitgaande van een significantieniveau van 0,10 is de kritieke waarde in de chi-kwadraatverdeling 2,71. Drie effecten blijken dan te verschillen naar vak. Op interconfessionele scholen gaan leerlingen significant meer achteruit in wiskunde dan in Nederlands. Op reformatorische scholen geldt juist het omgekeerde. Leerlingen gaan daar achteruit in Nederlands en vooruit in wiskunde. Het contrast tussen de beide effecten, zie ook Bijlage 5C, is -0,40 en de chi-kwadraat waarde ligt in het kritieke gebied. Het derde effect dat verschil naar vak is dat voor de invalregeling bij lesuitval. Voor Nederlands is dit effect sterker dan voor wiskunde.

In het spaarzame model is voor Nederlands de verklaarde variantie 6,1 procent op leerling-, 17,0 procent op klas- en 19,2 op schoolniveau. Voor wiskunde zijn deze percentages 10,7, 26,9 en 26,9 procent.

Tabel 5.8: Multiniveau analyse op leervorderingen. Variantie- en covariantiecomponenten en modelfit

	VARIANTIE NEDERLANDS	VARIANTIE WISKUNDE	COVARIANTIE	DEVIANCE)
INDIVIDUEEL	PAR. (S.E.)	PAR. (S.E.)	PAR. (S.E.)		
Leerlingniveau	0,855 (0,015)	0,826 (0,016)	0,091 (0,012)	33.987,6	
Klasniveau	0,059 (0,010)	0,064 (0,011)	0,016 (0,008)		
Schoolniveau	0,066 (0,013)	0,131 (0,022)	0,025 (0,013)		
SCHOOLTYPE					
Leerlingniveau	0,847 (0,015)	0,818 (0,016)	0,086 (0,012)	33.864,8	122,8
Klasniveau	0,052 (0,009)	0,064 (0,011)	0,023 (0,007)		(df=12)
Schoolniveau	0,083 (0,015)	0,122 (0,021)	0,038 (0,013)		p<0,001

RANDOM EFFECTEN					
Leerlingniveau	0,848 (0,014)	0,795 (0,016)	0,085 (0,012)	33.835,6	29,2
Klasniveau	0,052 (0,009)	0,041 (0,010)	0,021 (0,007)		(df=2)
Schoolniveau	0,083 (0,015)	0,114 (0,019)	0,037 (0,013)		p<0,001
CONTEXTUELE KENMERKEN					
Leerlingniveau	0,847 (0,015)	0,795 (0,016)	0,085 (0,012)	33.793,9	41,7
Klasniveau	0,051 (0,009)	0,041 (0,010)	0,020 (0,007)		(df=18)
Schoolniveau	0,065 (0,013)	0,093 (0,017)	0,034 (0,011)		p<0,001
SCHOOLKENMERKEN					
Leerlingniveau	0,845 (0,015)	0,795 (0,016)	0,085 (0,012)	33.727,7	66,2
Klasniveau	0,054 (0,009)	0,042 (0,010)	0,019 (0,007)		(df=27)
Schoolniveau	0,037 (0,009)	0,080 (0,015)	0,026 (0,009)		p<0,001
SPAARZAAM MODEL					
Leerlingniveau	0,845 (0,015)	0,794 (0,016)	0,085 (0,012)	33.760,6	32,9
Klasniveau	0,053 (0,009)	0,040 (0,010)	0,019 (0,007)		(df=23)
Schoolniveau	0,045 (0,010)	0,095 (0,017)	0,021 (0,010)		p=0,085

5.5 Covariantie- versus verschilscore-benadering

Uit een vergelijking van de covariantie-benadering en de verschilscore-benadering werd aan het begin van dit hoofdstuk opgemaakt dat beide methoden elkaar goed aanvullen. Afhankelijk van de vraag moet een onderzoeker een van de methoden kiezen. Wanneer een onderzoeker geïnteresseerd is in welke leerlingen op een bepaald moment een voorsprong of achterstand hebben in het onderwijs, dan is een covariantie-analyse gewenst. Wanneer een onderzoeker wil weten welke leerlingen voorop of achterop raken in het onderwijs, dan moet er worden gekozen voor een analyse van leerwinst of relatieve posities.

Uit het onderstaande overzicht blijkt dat de resultaten sterk verschillen tussen de modellen. Structurele kenmerken hebben een matige invloed op de leerprestaties van leerlingen. In de covariantie-benadering verklaren deze kenmerken respectievelijk 15 en 8 procent van de variantie in Nederlands en wiskunde. In de verschilscore-benadering hebben deze kenmerken (vrijwel) geen invloed. Dit betekent dat de invloed van structurele kenmerken vooral geldt in de periode voor en niet op de vorderingen in het voortgezet onderwijs. Ook de andere leerling- en gezinskenmerken, de culturele kenmerken, en het door de leerling gevolgde schooltype hebben in de verschilscore-benadering duidelijk een minder sterke invloed dan in de covariantiebenadering. Voor de contextuele en schoolkenmerken geldt dat de verschillen tussen beide benaderingen minder groot

zijn. Toch zijn de resultaten niet identiek. Een contextueel kenmerk als het percentage allochtone leerlingen hangt wel samen met leerprestaties, maar niet met leervorderingen. In grote lijn is de richting van effecten van schoolkenmerken in beide benaderingen gelijk en wisselt de sterkte van de effecten enigszins. Naast vaste effecten worden er in beide modellen ook random effecten gevonden. Regressie-effecten tussen klassen en scholen ontbreken in beide gevallen.

Wanneer wordt gekeken naar de intra-klasse correlatiecoëfficiënt dan valt op dat in de covariantie-benadering veel meer variantie is tussen klassen en scholen. De bruto variantie is op deze niveaus te zamen 51 procent voor Nederlands en 65 procent voor wiskunde. In de verschillscore-benadering is dit 13 en 19 procent. Dit grote verschil komt doordat leerprestaties sterk samenhangen met het samenstelling van klassen en scholen. Leervorderingen daarentegen worden veel minder door de compositie van scholen beïnvloed. De verschillscores variëren dus meer binnen dan tussen klassen en scholen.

Als gekeken wordt naar verschillen in netto variantie (waarbij gecontroleerd wordt voor leerling- en gezinskenmerken en het door leerlingen gevolgde schooltype), dan zijn de verschillen tussen beide benaderingen minder groot. Toch is er in de verschillscore-benadering nog steeds minder variatie tussen scholen dan in de covariantie-benadering. De variantie tussen scholen is ongeveer eenderde kleiner in de analyse van verschillen in relatieve posities dan in de covariantie-analyse. Voor wiskunde is er in plaats van 17 procent 12 procent residuele tussenschoolse variantie. Voor Nederlands is het 12 versus 9 procent.

In klassen en scholen waar goed wordt gepresteerd op Nederlands, zijn de prestaties op wiskunde ook goed. De correlaties tussen de afhankelijke variabelen is in de covariantie-benadering ongeveer 0,90. In de verschillscore-benadering is deze samenhang veel zwakker. Op leerlingniveau (binnen klassen binnen scholen) blijken de correlaties in beide benaderingen laag te zijn.

De totaal verklaarde variantie voor leerprestaties is hoog: 58 en 74 procent op leerlingniveau en ongeveer 90 procent op klas- en schoolniveau. In de verschillscore-benadering wordt minder variantie gebonden. De percentages zijn respectievelijk 6 en 11 procent op leerlingniveau en 17 tot 27 procent op klas- en schoolniveau. In paragraaf 5.3 is er al op gewezen dat hierbij rekening moet worden gehouden met de verschillen in betrouwbaarheid (zoals gemeten door Cronbachs alfa) van de afhankelijke variabelen in beide methoden. Samenvattend blijkt uit de vergelijking van de beide modellen dat de uitkomsten van de modellen verschillen. De covariantie- en verschillscore-benadering zijn duidelijk verschillende analysemethoden.

<i>Covariantie-benadering - leerprestaties -</i>	<i>Analytisch model</i>	<i>Verschilscore-benadering - leervorderingen -</i>
matige invloed sterke invloed sterke invloed zwakke invloed aanwezig afwezig	structurele kenmerken culturele kenmerken schooltype schoolkenmerken random effecten regressie-effecten tussen klassen en scholen	geen invloed zwakke invloed zwakke invloed zwakke invloed aanwezig afwezig
Ned / Wis 29 % / 38 % 22 % / 27 % 49 % / 35 %	bruto variantie: schoolniveau klasniveau leerlingniveau	Ned / Wis 6 % / 13 % 7 % / 6 % 87 % / 81 %
Ned / Wis 12 % / 17 % 6 % / 6 % 82 % / 77 %	netto variantie: schoolniveau klasniveau leerlingniveau	Ned / Wis 9 % / 12 % 5 % / 6 % 86 % / 82 %
0,93 0,90 0,33	bruto correlaties Nederlands - wiskunde: schoolniveau klasniveau leerlingniveau	0,23 0,28 0,13
Ned / Wis 89 % / 91 % 87 % / 92 % 58 % / 74 %	verklaarde variantie: schoolniveau klasniveau leerlingniveau	Ned / Wis 19 % / 27 % 17 % / 27 % 6 % / 11 %

5.6 Besluit

De covariantie- en verschilscore-benadering blijken duidelijk van elkaar te verschillen, wat komt doordat de twee analysemethoden op verschillende vragen een antwoord geven. Wanneer een onderzoeker geïnteresseerd is in welke leerlingen op een bepaald moment een voorsprong of achterstand hebben in het onderwijs, dan is een covariantie-analyse gewenst. Wanneer een onderzoeker wil weten welke leerlingen voorop of achterop raken in het onderwijs, dan moet er worden gekozen voor een analyse van leerwinst of relatieve posities. Hierbij wijs ik er nogmaals op dat het gebruik van longitudinale gegevens niet automatisch betekent dat een onderzoeker naar veranderingen kijkt. In een covariantie-analyse zijn er minimaal twee meetmomenten, maar heeft de analyse betrekking op een toestand op een bepaald moment. Beide methoden kunnen elkaar aanvullen, doordat ze elk antwoord geven op een andere vraag. Door de fundamentele verschillen tussen de methoden is de vraag of de ene methode beter is dan de andere minder interessant.

Op verschillen in relatieve posities blijken structurele kenmerken nauwelijks en culturele kenmerken veelal wel een effect te hebben. Van de structurele kenmerken is alleen de sekse van belang voor het verklaren van verschillen in relatieve posities. Van de culturele kenmerken zijn er effecten van de huiswerkcontrole (cirkel drie), het leefpatroon (op straat met vrienden, bijbaantje, uitgaan), het zelfbeeld (cirkel twee), de mening over docenten en regels, het aantal dagen huiswerk en de prestatiemotivatie (cirkel één). Behalve van deze leerling- en gezinskenmerken gaat er ook van het schooltype een effect uit.

Op klasniveau heb ik gekeken naar random effecten en regressie-effecten tussen klassen en scholen. In tegenstelling tot het laatste soort effecten blijken er wel random effecten. Er zijn namelijk random hellingen voor sekse en aantal dagen huiswerk. Het effect van deze leerlingkenmerken verschilt voor wiskunde van klas tot klas.

De uitkomsten tussen het covariantie-analytisch model en het variantie-analytisch model van verschillen in relatieve posities verschillen sterk. De etniciteit, het opleidingsniveau van vader en moeder, het aantal ouders (cirkel vijf), de autoritatieve opvoedingsstijl (cirkel vier), het tijd besteden aan lezen en hobby's (cirkel twee), de mening over klasgenoten en de non-verbale intelligentie (cirkel één) hebben wel een effect op de toetsscores Nederlands en wiskunde, maar geen effect op de verschillen in relatieve positie voor beide vakken. Hieruit kan worden opgemaakt dat de achterstand van leerlingen uit lagere sociale milieus of van allochtone leerlingen niet groter wordt in het voortgezet onderwijs. Omgekeerd is er in de analyse van verschillen in relatieve posities juist een effect van de prestatie-

motivatie. Voor de overige kenmerken geldt wel dat de richting van effecten voor beide modellen gelijk is, maar dat de sterkte van effecten opnieuw sterk verschilt.

Op klas- en schoolniveau blijken er in beide modellen geen regressie-effecten tussen klassen en scholen te zijn, maar wel random effecten en invloeden van enkele (contextuele) schoolkenmerken. Wat de invloed van contextuele kenmerken betreft blijkt dat van de locatiegrootte en de denominatie een invloed uitgaat op de verschillen in relatieve positie op beide vakken. Dit duidt erop dat deze kenmerken van invloed zijn op processen binnen de school. Grootschaligheid van onderwijs is voor de vorderingen van alle leerlingen negatief. Het is dus raadzaam om leerlingen zo veel mogelijk te spreiden over verschillende locaties in plaats van leerlingen allemaal op een locatie te huisvesten.

Daarentegen ontbreekt een effect van het percentage allochtone leerlingen in de analyse van verschillen in relatieve posities. Voor dit kenmerk geldt dat het weliswaar samenhangt met leerprestaties, maar dat er geen invloed van uitgaat op leervorderingen. Het percentage allochtone leerlingen is gerelateerd aan prestaties van leerlingen vanwege het instroomniveau en niet vanwege de processen binnen scholen. In de covariantie-benadering kunnen schoolkenmerken een effect hebben die niets te maken hebben met de kwaliteit van het onderwijsleerproces van de school maar die wel deels de verschillen in prestaties tussen scholen kunnen verklaren. In de verschilscore-benadering blijken zulke kenmerken geen effect te hebben.

Voor de overige schoolkenmerken is er ook een aantal effecten gevonden. Leerlingen boeken meer vooruitgang in Nederlands, wanneer scholen lesuitval opvangen door een invalregeling. De aard van de brugperiode resulteert ook in verschillen in vooruitgang. In scholen met een heterogene brugperiode boeken leerlingen, met uitzondering van vwo-leerlingen, minder vorderingen in Nederlands dan in scholen met een dakpansgewijze of homogene brugperiode. In scholen met een breed-horizontaal mentoraatsysteem boeken leerlingen minder vorderingen tussen leerjaar één en drie. Van de docent-leerling-ratio gaat in het beroepsonderwijs een effect uit op de leervorderingen.

Ook voor een analyse naar kenmerken van effectieve scholen maakt het uit welk model wordt gekozen. Kenmerken die betrekking hebben op het aanvangsniveau van leerlingen, zoals het percentage allochtone leerlingen, hebben soms wel een effect in de covariantie-benadering, maar geen effect in de verschilscore-benadering. Daarentegen zijn de uitkomsten voor kenmerken die betrekking hebben op processen in scholen meer overeenkomstig. Al met al blijkt uit een vergelijking van de covariantie- en de verschilscore-benadering dat de modellen op verschillende vragen antwoord geven. Door beide modellen te gebruiken, kan ant-

woord worden gegeven op de vraag naar zowel het niveau op een bepaald moment als de vooruitgang tussen twee momenten. Voor onderwijsonderzoek leidt dat tot een verrijking.

In de analyse van verschillen in relatieve posities zijn er op leerlingniveau minder effecten dan in hoofdstuk 3. Toch zijn er verscheidene kenmerken die ook in deze benadering van invloed zijn. Wanneer de modellen op verklaarde variantie worden vergeleken, dan is echter het contrast schril. Op leerlingniveau wordt in de analyse van verschillen in relatieve posities minder dan 2,5 procent variantie verklaard. Wanneer het effect van de eerdere prestaties niet wordt meegerekend, bedraagt dit percentage voor de covariantie-benadering 24 procent. De lage verklaarde variantie in de verschilscore-benadering heeft deels te maken met de grote restvariantie op het individuele niveau, maar impliceert ook dat nauwelijks bekend is waarom bepaalde leerlingen in de onderbouw een voorsprong boeken en anderen achterstand oplopen. Om vooruitgang in prestaties te verklaren, volstaan de gebruikelijke verklaringen uit onderzoek naar leerprestaties slechts in geringe mate. Over de invloeden op onderwijsprestaties in de onderbouw van het voortgezet onderwijs is dan ook nog te weinig bekend.

Eindnoten

ⁱ De random helling van rekenen, zie het vorige hoofdstuk, heeft bijvoorbeeld invloed op de variantie in wiskunde.

ⁱⁱ Wanneer leerlingen in het voortgezet onderwijs te maken hebben gehad met ingrijpende veranderingen in de structurele kenmerken, zoals een ouder die werkloos wordt of het uit elkaar gaan van ouders, dan zou er een effect op leervorderingen te verwachten zijn.

ⁱⁱⁱ Voor Nederlands is de ware variabiliteit overigens nog groter, omdat de correlatie tussen begin- en eindmeting beduidend minder sterk is.

^{iv} Ten derde heb ik naar regressie-effecten tussen klassen en scholen gekeken. Net als in het vorige hoofdstuk heb ik op klas- en schoolniveau dergelijke effecten niet gevonden. De effecten van leerling- en gezinskenmerken zijn dus effecten binnen klassen en niet tussen klassen of scholen.

^v Voordat de integrale analyses zijn uitgevoerd, is eerst per schooltype gekeken naar de effecten van contextuele en schoolkenmerken. Zie Bijlage 4B en 4C voor soortgelijke analyses.

^{vi} Hierbij moet overigens worden opgetekend dat teamstabiliteit en prestaties als extra toelatingscriterium wel een effect hebben op wiskundevorderingen, zie Bijlage 5C. In het spaarzame model kunnen de effecten van deze twee kenmerken echter worden weggelaten zonder dat het model significant verslechtert.

6 | Conclusie

6.1 Leerlingen – klassen - scholen

Met gegevens van ongeveer 7.000 leerlingen afkomstig uit 450 klassen van 150 scholen voor voortgezet onderwijs ben ik nagegaan in hoeverre verschillen in prestaties en vorderingen Nederlands en wiskunde kunnen worden verklaard door kenmerken van leerlingen, gezinnen, docenten en scholen. Om deze vraag te beantwoorden heb ik rekening gehouden met de geneste structuur van de gegevens. Hierbij is gebruik gemaakt van multiniveau analyse. De informatie komt uit het cohortonderzoek VOCL'93. Deze dataset bevat gegevens van leerlingen die in het schooljaar 1993-'94 op het ivbo, het vbo, de mavo, de havo of het vwo zaten. Voor de leerlingen die in de analyses zijn meegenomen geldt dat ze in het schooljaar 1995-'96 in leerjaar drie zaten. De analyses zijn dus gebaseerd op leerlingen die onvertraagd zijn doorgestroomd.

Een geneste structuur

Multiniveau analyse heeft verscheidene voordelen ten opzichte van analyses op één niveau, zoals een regressieanalyse met geaggregeerde of gedesaggregeerde gegevens. Door rekening te houden met de hiërarchie van gegevens (leerlingen binnen klassen binnen scholen) zijn de uitkomsten correcter en beter interpreteerbaar dan die van een analyse op één niveau. Een ander voordeel van multiniveau analyse is dat er naar differentiële effecten kan worden gekeken. Zo blijkt het effect van etniciteit op prestaties niet in alle klassen hetzelfde (paragraaf 4.4.2). In sommige klassen is dit effect sterker en in andere klassen zwakker, wat inhoudt dat voor allochtone leerlingen de ene klas gemiddeld genomen gunstiger en de andere klas ongunstiger is. Een derde voordeel is dat er met multiniveau analyse een onderscheid kan worden gemaakt tussen binnenschoolse en tussenschoolse regressie. Na uitvoerige controle op individueel niveau bleken er in dit onderzoek geen regressie-effecten tussen klassen of scholen te zijn. Het percentage allochtone leerlingen kan overigens wel als een voorbeeld van een tussenschoolse regressie-effect worden beschouwd (zie paragraaf 4.5). Niet alleen op individueel maar ook op schoolniveau blijkt er namelijk een effect van etniciteit te zijn. Onder constanthouding van de kenmerken van leerlingen en gezinnen, scoren leerlingen op scholen met veel allochtonen slechter dan scholen met weinig allochtonen. Van de

compositie van de school gaat dus een effect uit. Verder geldt dat met multiniveau analyse scholen gemakkelijk te rangordenen zijn door gebruik te maken van zogeheten ‘posterior means’. In de multiniveau analyse in hoofdstuk 2 zijn met deze schattingen rapportcijfers van scholen verkregen.

Met het Trouw-onderzoek naar schoolprestaties heb ik geïllustreerd dat de schattingen van de netto prestaties van scholen sterk verschillen voor analyses op één dan wel twee niveaus. Voor meer dan de helft van de scholen geldt dat de rapportcijfers van scholen één of meer punten verschillen (op een schaal van vier tot acht) naar gelang de keuze van het aantal niveaus. Wanneer er in onderzoek geen rekenschap wordt gegeven van de aanwezigheid van een geneste structuur, dan is bij de interpretatie van de uitkomsten op zijn minst voorzichtigheid geboden. Succesmaten uit een analyse op één niveau, zoals de rapportcijfers uit het Trouw-onderzoek, geven dan ook zeker geen betrouwbaar beeld van de prestatieverschillen tussen scholen.

Doorgaans leiden analyses op één niveau tot fouten, wanneer de structuur genest is. Vandaar dat een onderzoeker er niet aan ontkomt om een multiniveau analyse te verrichten. In de discussie over openbare schoolgegevens van scholen komen Bosker en Scheerens (1999) tot dezelfde conclusie. Waar het om draait bij het bepalen van de door de school toegevoegde waarde (*posterior means*), zo stellen zij, is de vaststelling van de gemiddelde mate waarin individuele leerlingen boven of onder de statistische verwachting presteren op een school. Dat is iets anders dan de vaststelling of een schoolpopulatie als geheel verschilt van de verwachte score.

In hoofdstuk 2 heb ik ook laten zien dat de uitkomsten eveneens sterk verschillen tussen een analyse met bruto of netto prestaties. Een indicator voor de bruto prestatie van een school is bijvoorbeeld het gemiddelde niveau dat leerlingen op Nederlands en wiskunde hebben. Om op eerlijke wijze zichtbaar te maken ‘wat scholen er toe doen’ is het echter nodig om rekening te houden met de omstandigheden waarin scholen verkeren. Er moeten dus netto in plaats van bruto prestaties worden gehanteerd. De netto prestatie van een school is dan bijvoorbeeld het gemiddelde niveau van leerlingen op Nederlands en wiskunde, nadat er gecontroleerd is voor verschillen in de samenstelling van leerlingenpopulaties. Met gegevens op één niveau, zoals het in Trouw-onderzoek, blijkt het slechts ten dele mogelijk om de netto prestatie van een school vast te stellen. Wanneer er echter gegevens op zowel individueel als schoolniveau worden verzameld, dan kan de netto prestatie van een school in een multiniveau analyse worden bepaald.

Uit de multiniveau analyses in hoofdstuk 3 en 4 blijkt dat in Nederland de bruto variantie tussen klassen en scholen hoog is. In dit onderzoek is deze variantie 51

procent voor Nederlands en 65 procent voor wiskunde. De variantie tussen klassen en scholen is zo hoog, omdat in het Nederlandse voortgezet onderwijs leerlingen worden opgedeeld naar schooltype. In *Education at a glance* (OECD, 1998) blijkt dat Nederland wat relatieve tussenschoolse variantie in het voortgezet onderwijs betreft internationaal koploper is. Landen die hier ook hoog op scoren zijn Ierland, Duitsland, Zwitserland, Oostenrijk en het Vlaamse deel van België. In Korea en Japan is de bruto variantie tussen klassen en scholen daarentegen minder dan 5 procent. Deze maat zegt overigens niets over de ongelijkheid in prestaties binnen een land. Tussen de tussenschoolse variantie (een relatieve maat) en de algehele variatie in prestaties (een absolute maat) blijkt geen duidelijke relatie te bestaan (OECD, 1998).

Het individuele niveau: prestaties

Voor een verklaring van verschillen in prestaties zijn de gegevens op individueel niveau het belangrijkste. Het niveau waarop leerlingen presteren op Nederlands en wiskunde wordt bepaald door zowel structurele als culturele kenmerken. Zelfs als er rekening wordt gehouden met de eerdere prestaties van leerlingen, is het om na te gaan welke leerlingen een voorsprong of achterstand hebben van belang om te weten wat de *opleiding van hun ouders* is en of ze *autochtoon of allochtoon* zijn. Wanneer het effect van deze kenmerken in samenhang wordt bekeken, dan is er sprake van een sterk effect. Een autochtone leerling met ouders met een hogere beroepsopleiding komt bijvoorbeeld een kwart standaarddeviatie hoger uit dan een allochtone leerling met ouders met alleen basisonderwijs. Ter vergelijking, het verschil in prestaties tussen leerlingen die een niveau in schooltype van elkaar verschillen, bijvoorbeeld een mavo- en een havo-leerling, is ongeveer een halve standaarddeviatie. Het verschil tussen allochtonen en autochtonen valt niet weg, wanneer er rekening wordt gehouden met de kenmerken van de sociaal-economische status. Etniciteit heeft dus een extra effect op prestaties en dan niet alleen op Nederlands maar ook op wiskunde.

Onder constant houding van de overige kenmerken heeft de *seks* eveneens een sterk effect op de prestaties van leerlingen. Meisjes blijken gemiddeld genomen beter te scoren op Nederlands en jongens op wiskunde.¹ Van het opgroeien in een *tweeoudergezin* gaat een positief effect uit op de wiskundeprestaties.

Naast effecten van structurele kenmerken zijn er ook effecten van culturele kenmerken. Van de interactie tussen ouders en leerlingen hebben de *huiswerkcontrole* en de *autoritatieve opvoedingsstijl* een behoorlijke samenhang met de prestaties van leerlingen. Leerlingen die niet door hun ouders worden aangespoord tot en gecontroleerd op het maken van huiswerk en die de helft van de beslissingen na overleg met hun ouders nemen (er wordt dus geregeld gebruik gemaakt van de au-

toritatieve opvoedingsstijl) scoren voor Nederlands 0,33 standaarddeviatie en voor wiskunde 0,18 standaarddeviatie hoger dan leerlingen die geregeld te maken hebben met huiswerkcontrole en die niet autoritatief worden opgevoed.

Van de algemene kenmerken van leerlingen heeft het *zelfbeeld* van leerlingen een sterk positief effect. Dit effect is net zo groot als het gezamenlijke effect van het opleidingsniveau van beide ouders. Ook het *leefpatroon* heeft een sterke invloed op de prestaties van leerlingen. Leerlingen die de helft van hun vrije tijd besteden aan uitgaan, een bijbaantje of op straat met vrienden rondhangen en die geen tijd besteden aan lezen of hobby's scoren een kwart standaarddeviatie lager op Nederlands en een vijfde standaarddeviatie lager op wiskunde dan leerlingen die 20 procent van hun vrije tijd besteden aan lezen of hobby's en die geen tijd besteden aan de drie andere aspecten.

De grootste effecten op de prestaties komen natuurlijk van de effecten van de *eerdere prestaties*. Wanneer een leerling bijvoorbeeld op alledrie de toetsen (taal, informatieverwerking en rekenen) één standaarddeviatie boven het gemiddelde scoort, presteert een leerling twee jaar later op havo-niveau. Bij twee standaarddeviaties boven het gemiddelde scoort die leerling in leerjaar drie op vwo-niveau. Een verschil in één standaarddeviatie resulteert in een verschil van ongeveer één schooltype. De eerdere prestaties zijn daarmee een goede voorspeller van de latere prestaties van leerlingen. Voor de andere kenmerken uit cirkel één geldt dat het aantal dagen huiswerk een behoorlijke invloed heeft. Wanneer een leerling in plaats van vijf maar twee dagen in de week huiswerk maakt resulteert dat in een afname van de prestaties op beide toetsen met eentiende standaarddeviatie.

Wanneer de verschillende cirkels met elkaar worden vergeleken, dan blijkt dat de prestaties van leerlingen meer worden verklaard door kenmerken van leerlingen dan van gezinnen. Uit het literatuuroverzicht komen veel meer kenmerken van leerlingen dan van de gezinsinteractie naar voren. Ook blijkt dat de verklaarde variantie veel sterker toeneemt door de opname van leerlingkenmerken (cirkel 1' en 2) dan door gezinskenmerken (cirkel 3 en 4). Voor leerlingkenmerken geldt verder dat de prestaties van leerlingen meer door schoolse dan door algemene kenmerken worden verklaard. Deze conclusie kan voor gezinskenmerken niet worden getrokken. Voor de verklaring van verschillen in prestaties blijken culturele kenmerken verder belangrijker dan structurele kenmerken. De structurele kenmerken binden minder dan 10 procent van de variantie. De culturele kenmerken verklaren vier à vijf keer zo veel variantie.

Het *schooltype* heeft ook een zelfstandig effect op de leerprestaties. Dit kenmerk geeft een indicatie van zowel het prestatieniveau van een leerling als de gelegenheid tot leren. Naarmate het schooltype hoger is, krijgen leerlingen een moeilijkere leerstof aangeboden, waardoor bijvoorbeeld vwo-leerlingen meer ge-

legenheid tot leren hebben dan overige leerlingen. Het schooltype heeft niet alleen een extra effect. Het blijkt ook andere effecten te mediëren, zoals het effect van het opleidingsniveau van ouders of het aantal dagen waarop door leerlingen huiswerk wordt gemaakt.

Het individuele niveau: vorderingen

In het voorgaande stond de vraag centraal in hoeverre de prestaties van leerlingen op één bepaald tijdstip kunnen worden verklaard uit allerlei kenmerken van leerlingen en gezinnen. Om de netto prestaties van leerlingen te bepalen wordt hierbij rekening gehouden met de prestaties van leerlingen op een eerder tijdstip. De toetsscores in leerjaar drie zijn de afhankelijke variabelen en de eerdere prestaties in leerjaar één behoren tot de onafhankelijke variabelen. Een dergelijk model is standaard in het schooleffectiviteitsonderzoek en staat bekend als het *covariantie-analytisch model*.

Met het covariantie-analytisch model kunnen prestaties van leerlingen op een bepaald tijdstip worden verklaard. In deze verklaring gaat het om de vraag welke leerlingen een achterstand of voorsprong *hebben*. Vervolgens kan dan worden nagegaan welke kenmerken invloed hebben op de *leerprestaties*. Met dit model wordt dus naar invloeden op leerprestaties en niet naar invloeden op leervorderingen gekeken. Wanneer een onderzoek op dit model is gebaseerd, kan dan ook niet over effecten op vooruitgang in prestaties worden gesproken. Om naar vooruitgang in prestaties te kijken, moet worden gekozen voor een *variantie-analytisch model van verschillen in relatieve posities*. De afhankelijke variabelen hierbij zijn bijvoorbeeld de *verschilsscores* tussen de gestandaardiseerde prestaties in leerjaar drie en één.

De uitkomsten van de covariantie-benadering en de verschilscore-benadering blijken op individueel niveau sterk van elkaar af te wijken. Structurele kenmerken (ethniciteit, opleidingsniveau van vader en moeder of aantal ouders) en een deel van de culturele kenmerken (autoritatieve opvoedingsstijl, tijd besteden aan lezen en hobby's, mening over klasgenoten en non-verbale intelligentie) hebben wel een effect op de toetsscores Nederlands en wiskunde, maar geen effect op de verschillen in relatieve positie voor beide vakken. Omgekeerd is er in de analyse van verschillen in relatieve posities juist een effect van een cultureel kenmerk als de prestatiemotivatie, terwijl dat effect in de covariantie-benadering ontbreekt. Behalve van deze leerling- en gezinskenmerken gaat er ook van het schooltype een zwakke invloed uit op de vorderingen van leerlingen.

Uit de analyses blijkt dat op de relatieve positie op beide vakken structurele kenmerken nauwelijks een effect hebben. Afwijkend van wat vaak wordt aangenomen, bijvoorbeeld Hacquebord (1989), wordt de achterstand van allochtone

leerlingen niet groter in het voortgezet onderwijs. Daarentegen hebben verscheidene culturele kenmerken wel een effect op de vorderingen Nederlands en wiskunde. De veranderingen in prestaties komen door kenmerken van leerlingen, zoals hun inzet voor school of hun leefpatroon.ⁱⁱ Vanuit meritocratisch oogpunt is dit een gunstige uitkomst. Kenmerken die gegeven zijn, zoals de sociaal-economische status en de etniciteit, hebben geen invloed op verschillen in relatieve posities, maar culturele kenmerken wel.

Toch moet de invloed van deze veranderbare kenmerken niet worden overschat. De culturele kenmerken verklaren namelijk weinig van de variantie in verschillen in relatieve posities. Individuele kenmerken hebben dus weinig invloed op leerwinst. Bij de intrede in het voortgezet onderwijs nemen leerlingen een bepaalde positie qua prestaties in. In die posities treden tijdens de eerste drie leerjaren van het voortgezet onderwijs alleen in beperkte mate veranderingen op. Ter indicatie, de correlatie tussen de scores op de toetsen uit leerjaar één en drie is 0,55 voor Nederlands en 0,74 voor wiskunde.

Het klas- en schoolniveau: prestaties en vorderingen

Op klas- en schoolniveau zijn de overeenkomsten tussen de covariantie- en verschillscore-benadering sterker. De kenmerken die van invloed zijn op de prestaties in leerjaar drie en op de vorderingen tussen leerjaar één en drie zijn in grote lijn vergelijkbaar.

Voor beide benaderingen geldt dat na controle voor kenmerken op individueel niveau scholen in geringe mate van elkaar verschillen. De netto variantie in prestaties (waarbij gecontroleerd wordt voor leerling- en gezinskenmerken en het door leerlingen gevolgde schooltype) tussen scholen is 12 procent voor Nederlands en 19 procent voor wiskunde. De netto variantie op schoolniveau in de verschillscore-benadering is 9 procent voor Nederlands en 12 procent voor wiskunde. In de verschillscore-benadering geldt ook voor de bruto variantie dat slechts een klein deel van de variantie te maken heeft met systematische verschillen tussen scholen. De relatieve tussenschoolse variantie is 6 en 13 procent voor respectievelijk Nederlands en wiskunde. In de covariantie-benadering zijn de percentages veel hoger, namelijk 29 en 38 procent (tussen klassen en scholen zelfs 51 en 65 procent). Leerprestaties hangen namelijk sterk samen met de samenstelling van scholen, terwijl leervorderingen veel minder door de compositie van een school worden beïnvloed. Verder geldt in beide benaderingen dat er meer relatieve tussenschoolse variantie is voor wiskunde dan voor Nederlands, wat komt doordat prestaties en vorderingen op wiskunde meer samenhangen met wat er op school gebeurt dan die op Nederlands.

Van de *contextuele kenmerken* blijken het percentage allochtone leerlingen en de locatiegrootte beide invloed te hebben op de *prestaties* van leerlingen. Voor leerlingen in alle schooltypen geldt dat het percentage allochtone leerlingen een negatief effect heeft op hun wiskundeprestaties. Door een toename van een standaarddeviatie in het percentage allochtone leerlingen, gaat de score op de toets met 0,03 standaarddeviatie achteruit. Voor leerlingen die op de havo of het vwo zitten geldt dat ze niet alleen op wiskunde maar ook op Nederlands slechter presteren naarmate het percentage allochtone leerlingen op school stijgt. Voor elke standaarddeviatie verschil in het percentage allochtone leerlingen, verandert de score op de toets met 0,09 standaarddeviatie. Het percentage allochtone leerlingen hangt samen met talrijke andere kenmerken van de leerlingencompositie. Op zwarte scholen is het opleidingsniveau van ouders en het instroomniveau van leerlingen lager dan op witte scholen. In dit onderzoek blijkt het resultaat van dergelijke verschillen te zijn dat de prestaties van leerlingen lager zijn naarmate er meer allochtone leerlingen op een school zitten. Voor vervolgonderzoek zou het interessant zijn om na te gaan waarom het percentage allochtone leerlingen een negatief effect heeft op prestaties van leerlingen, en waarom de prestatie-achterstand voor havo/vwo-leerlingen niet alleen voor wiskunde maar ook voor Nederlands geldt. Zijn die prestaties lager, omdat het percentage allochtone leerlingen een meting is van ongemeten leerlingkenmerken? Of passen docenten op scholen met veel allochtone leerlingen hun lessen aan en verlagen ze hun standaarden? En waarom zijn er verschillen in de effecten voor Nederlands en wiskunde?

Voor leerlingen in het beroepsonderwijs maakt de locatiegrootte een verschil uit voor de prestaties op beide vakken. Wanneer het aantal leerlingen op een zelfde locatie toeneemt van 300 tot 600 of van 600 tot 1200, dan leidt dit tot een afname in prestaties van 0,07 standaarddeviatie op beide vakken. Aangezien vele leerlingen de gevolgen merken van een dergelijk schooleffect, moet dit effect als relevant worden beschouwd. Voor mavo-, havo- en vwo-leerlingen is er geen effect van de locatiegrootte op de prestaties.

In onderzoek buiten Nederland worden geregeld verschillen in prestaties aangetoond tussen openbaar en bijzonder onderwijs. Die verschillen zijn er dan voor zowel bruto als netto prestaties. Uit dit onderzoek blijkt dat er in Nederland verschillen in netto prestaties tussen denominaties zijn. In tegenstelling tot bijvoorbeeld Amerikaans onderzoek blijken openbare scholen niet slechter te presteren dan bijzondere scholen. Binnen de bijzondere sector blijken er met name verschillen te zijn. Hierbij valt vooral de achterstand van leerlingen op interconfessionele scholen op.

Op de *vorderingen* van leerlingen blijkt er van het percentage allochtone leerlingen geen en van de locatiegrootte en de denominatie wel een effect uit te gaan.

Het percentage allochtone leerlingen is, zoals eerder aangegeven, wel gerelateerd aan prestaties van leerlingen in leerjaar drie van het voortgezet onderwijs. Deze relatie hangt waarschijnlijk vooral samen met de achterstand die leerlingen op scholen met veel allochtone leerlingen hadden bij aanvang van het voortgezet onderwijs en niet met de processen in de onderbouw van het voortgezet onderwijs. De achterstand van leerlingen op 'zwarte scholen' wordt namelijk niet groter tussen leerjaar één en drie.

In de covariantie-benadering heeft de locatiegrootte alleen een negatief effect op de prestaties van leerlingen in het beroepsonderwijs. In de verschillscore-benadering wordt echter een negatief effect voor alle leerlingen gevonden. Dit negatieve effect duidt erop dat de processen op scholen verschillen naar gelang de locatiegrootte. Een effect van de denominatie is ook aanwezig. Opnieuw zijn er geen significante verschillen tussen openbare, protestants-christelijke en rooms-katholieke scholen. Wel valt op dat interconfessionele scholen opnieuw negatief naar voren komen. Leerlingen op dit type scholen blijken een wiskunde-achterstand op te lopen.

Van de *schoolkenmerken* zijn er enkele significante effecten. De teamstabiliteit blijkt een zwak positief effect te hebben op de leerprestaties op beide vakken. Door elk jaar dat het schoolteam langer bijeen is, blijken de prestaties van leerlingen met eentwintigste standaarddeviatie vooruit te gaan. Op de leervorderingen heeft dit kenmerk geen effect.

Wanneer scholen lesuitval opvangen door een invalregeling of een verplichte huiswerkles, dan blijken vbo/mavo-leerlingen op Nederlands een vijfde tot een derde standaarddeviatie hoger te presteren dan wanneer het uur komt te vervallen. Een invalregeling blijkt ook op de vorderingen van vbo/mavo-leerlingen een effect te hebben, maar het effect van een verplichte huiswerkles blijkt in de verschillscore-benadering niet significant.

Met uitzondering van vwo-ers, scoren leerlingen in heterogene brugklassen in zowel de covariantie- als de verschillscore-benadering lager op Nederlands dan leerlingen in een dakpansgewijze of homogene brugklas. Wanneer leerlingen van drie of meer schooltypen bij elkaar zaten in het eerste leerjaar van het voortgezet onderwijs, dan hebben die leerlingen in het derde leerjaar een achterstand opgelopen voor Nederlands. In de praktijk groeperen scholen leerlingen overigens steeds minder heterogeen. Meer dan de helft van de leerlingen zit in een homogene brugklas. Deze leerlingen hebben hun keuze voor een schooltype dus al voor het voortgezet onderwijs moeten maken. In een heterogene brugklas zit één op de tien leerlingen. De overige leerlingen zitten in een tussenvorm, een dakpansgewijze brugklas (leerlingen van twee schooltypen in dezelfde klas).

Wat de vormgeving van het mentoraatsysteem betreft valt op dat het verband in beide benaderingen er als volgt uit ziet: hoe uitgebreider dat systeem is opgezet, hoe slechter de prestaties van leerlingen. Alle vormen scoren lager dan een smal-verticaal mentoraatsysteem, de referentiegroep. Tot slot, blijkt dat (i)vbo-leerlingen slechter presteren naarmate de docent-leerling-ratio hoger is. Hier geldt dat vbo-leerlingen net als ivbo-leerlingen baat hebben bij een structuur waarin ze niet al te veel verschillende docenten hebben. In de verschilscore-benadering wordt dit effect alleen voor Nederlands gevonden. In het algemeen bleek in deze analyses dat het niet of nauwelijks mogelijk is om de invloed van contextuele kenmerken te verklaren door schoolkenmerken.

6.2 Wetenschappelijke relevantie

De conclusie dat leerling- en gezinskenmerken bepalend zijn voor verschillen in prestaties is in lijn met onderzoek van Coleman e.a. (1966) en Jencks e.a. (1972). In dertig jaar tijd is er in dit opzicht niet veel veranderd. Technisch gezien is er echter vooruitgang geboekt door de ontwikkeling van *multiniveau analyse*. Zoals aan het begin van dit hoofdstuk vermeld heeft multiniveau analyse verscheidene voordelen boven een één-niveau analyse van hiërarchische structuren.

In onderwijsonderzoek is het nog niet gebruikelijk om verscheidene afhankelijke variabelen simultaan in een *multivariaat model* op te nemen. In dit onderzoek zijn de analyses tegelijkertijd voor Nederlands en wiskunde verricht. Een voordeel hiervan is dat er kan worden getoetst of onafhankelijke variabelen een zelfde effect hebben op de twee afhankelijke variabelen. In de tweede plaats kunnen de correlaties tussen de afhankelijke variabelen op verschillende niveaus worden bepaald. Ook geldt dat leerlingen met een ontbrekende waarde op een van de afhankelijke variabelen in de analyse kunnen worden meegenomen, waardoor incomplete datasets efficiënter kunnen worden gebruikt. Een multivariaat multiniveau model heeft dus belangrijke voordelen boven de modellen die enkele decennia terug werden gehanteerd. Door de ontwikkeling van deze modellen kan er dus beter en uitgebreider onderzoek worden gedaan naar verschillen in leerprestaties en leervorderingen tussen leerlingen.

Door de ontwikkeling van multiniveau analyse is het goed mogelijk om naar de invloed van leerlingen, gezinnen, docenten en scholen in *samenhang* te kijken. Uit dit onderzoek blijkt dat het belangrijk is om op individueel niveau uitgebreid te controleren. In studies naar effectieve docenten en scholen blijkt die controle vaak beperkt. Met structurele kenmerken en eerdere prestaties van leerlingen houden effectiviteitsonderzoekers wel rekening, maar met culturele kenmerken niet.

Voor de verklaring van verschillen in prestaties tussen leerlingen zijn deze culturele kenmerken echter minstens zo belangrijk als structurele kenmerken. Voor de verklaring van toe- of afname in relatieve positie geldt zelfs dat structurele kenmerken niet of nauwelijks, maar culturele kenmerken wel van belang zijn. Om over effecten op klas- en schoolniveau uitspraak te doen moet daarom op individueel niveau met zowel structurele als culturele kenmerken rekening worden gehouden.

In onderzoek met verscheidene kenmerken op verschillende niveaus is het gevaar dat het onderzoek te sterk door de empirie wordt gestuurd. Met een sleepnet trekt de onderzoeker alle gegevens voort. Na verschillende analyses blijven er dan enkele kenmerken hangen en die zijn blijkbaar van invloed op de afhankelijke variabele. Om dat gevaar zo veel mogelijk te ontlopen is er in dit onderzoek een *structuur van concentrische cirkels* gehanteerd. Op grond van theoretische overwegingen zijn er kenmerken in deze structuur opgenomen. Met deze structuur zijn vervolgens cirkel voor cirkel analyses uitgevoerd. Met het gebruik van deze structuur kunnen de verschillende invloeden goed worden geordend en kan ook inzicht worden gegeven in de (causale) relaties tussen de kenmerken onderling. Zo blijken de effecten van kenmerken als de opleiding van de moeder of de etniciteit voor een belangrijk deel te verklaren door meer direct aan prestaties gerelateerde kenmerken. Beide effecten worden voor ongeveer drievijfde verklaard door culturele kenmerken. Door de toevoeging van het schooltype worden de effecten van de opleiding van de moeder en van de vader zelfs volledig verklaard. Voor de *verklaring* van verschillen in prestaties op individueel niveau lijkt de structuur van concentrische cirkels dan ook bruikbaar.

Op klas- en schoolniveau beantwoorden de analyses gedeeltelijk aan de verwachtingen vooraf. Op beide niveaus zijn er op voorhand verscheidene kenmerken die voor verschillen in prestaties in de onderbouw van het Nederlandse voortgezet onderwijs van belang kunnen worden geacht. Deze verwachtingen worden voor een deel empirisch bevestigd. Voor verscheidene kenmerken, zeker voor die van docenten, geldt echter dat de beoogde meetinstrumenten uit VOCL'93 niet geschikt bleken. Over de kenmerken van effectieve docenten kan hier dan ook niets worden gezegd.

In het multiniveau model is alleen maar rekening gehouden met de brugklas waarin de leerling zich bevond en niet met de klas waarin de leerling in het derde leerjaar zat. Technisch is het mogelijk om een geneste structuur op te stellen, waarbij leerlingen zich in een bepaalde klas op tijdstip 1 en een andere klas op tijdstip 2 bevinden. Daartoe kan gebruik worden gemaakt van cross-classificaties (Goldstein, 1995²). Praktisch lijkt het echter niet aantrekkelijk om die structuur voor alle 7.000 leerlingen uit dit onderzoek te gebruiken. Voor sommige scholen

geldt dat leerlingen na de brugklas uiteen zwermen over tal van klassen, waardoor het aantal leerlingen per klas drastisch verminderde. Voor andere scholen geldt dat leerlingen bij elkaar in de klas blijven, waardoor een onderscheid tussen de variantie van het eerste en het derde leerjaar niet kan worden gemaakt. Voor een deel van de scholen uit VOCL'93 is het wellicht mogelijk om een analyse met cross-classificaties uit te voeren.

In de analyses zijn (i)vbo-, mavo-, havo- en vwo-leerlingen gecombineerd meegenomen. Het blijkt technisch goed mogelijk om de analyses voor *alle schooltypen tegelijkertijd* te verrichten. Hiervan is het voordeel dat op klas- en schoolniveau het aantal onderzoekselementen niet onnodig wordt verkleind. Door het maken van interacties kunnen bovendien effecten worden opgenomen die specifiek voor een bepaald schooltype gelden. Een onderzoek naar kenmerken van effectieve docenten en scholen hoeft dus niet per schooltype te geschieden.

Tot slot geldt dat er met name op individueel niveau verschillen zijn tussen de uitkomsten van *de covariantie- en de vershilscore-benadering*. Een covariantie-analyse geeft geen antwoord op de vraag welke leerlingen leerwinst boeken. Alleen met een analyse van leerwinst of relatieve posities kan er worden gekeken naar de voor- of achteruitgang in prestaties van leerlingen. Uit een vergelijking van de twee analysemethoden blijkt dat ze elkaar goed aanvullen. Afhankelijk van de vraag moet een onderzoeker een van de methoden kiezen. Wanneer een onderzoeker geïnteresseerd is in welke leerlingen op een bepaald moment een voor- sprong of achterstand hebben in het onderwijs, dan is een covariantie-analyse gewenst. Wanneer een onderzoeker wil weten welke leerlingen voorop of achterop raken in het onderwijs, dan moet er worden gekozen voor een analyse van leerwinst of relatieve posities. In tegenstelling tot een onderzoek naar kenmerken van effectieve leerlingen en gezinnen, blijkt het voor een onderzoek naar kenmerken van effectieve scholen minder uit te maken welk model wordt gekozen.

Technisch en inhoudelijk is er de afgelopen decennia dus vooruitgang geboekt in het onderzoek naar verschillen in prestaties. Het antwoord op de vraag naar *verschillen in prestaties* kan daarom met meer kracht en met meer nuance worden gegeven. Voor een verklaring van verschillen in prestaties zijn kenmerken op individueel niveau cruciaal. Van deze kenmerken is de invloed van culturele kenmerken veel groter dan van structurele kenmerken. Van de culturele kenmerken is de invloed van de schoolse leerlingkenmerken het grootst, gevolgd door de algemene leerlingkenmerken en de gezinskenmerken.

Voor een verklaring van *verschillen in relatieve posities* geldt dat kenmerken op individueel niveau een veel minder goede verklaring vormen. Met uitzondering van de sekse, hebben structurele kenmerken geen invloed op verschillen in relatieve posities. Verscheidene culturele kenmerken hebben wel een invloed op die

verschillen, waarbij geldt dat de effecten van leerlingkenmerken sterker zijn dan die van gezinskenmerken. Tussen algemene en schoolse kenmerken is geen duidelijk verschil in de sterkte van effecten. Wat verklaarde variantie betreft valt op dat de kenmerken op individueel niveau nauwelijks variantie binden. Hieruit kan worden opgemaakt dat nauwelijks bekend is waarom bepaalde leerlingen in de onderbouw van het voortgezet onderwijs voorsprong boeken en anderen achterstand oplopen.

6.3 Maatschappelijke relevantie

Doelmatigheidsprikkels

Het ministerie van Onderwijs, Cultuur en Wetenschappen wil dat scholen zich publiekelijk gaan verantwoorden voor hun prestaties tegenover ouders en de overheid. Het ministerie beoogt op die manier ouders beter voor te lichten, de resultaten van scholen vergelijkbaar te maken en scholen te prikkelen tot betere prestaties.

In het voortgezet onderwijs heeft dit geleid tot de ontwikkeling van kwaliteitskaarten. Bosker en Scheerens (1999) beargumenteren dat het publiceren van opbrengstgegevens van scholen in de praktijk slechts een geringe invloed zal hebben. De rapportcijfers van Trouw suggereren, in tegenspraak met de uitkomsten uit onderwijsonderzoek, een enorme spreiding in kwaliteit tussen scholen. Deze spreiding is echter een artefact van de methode van het Trouw-onderzoek. Wanneer scores worden gestandaardiseerd met een zes als gemiddelde en een standaarddeviatie van één dan heeft automatisch 40 procent van de scholen een score tussen 5,5 en 6,5. Deze scholen krijgen dus een zes als cijfer. Van de overige scholen scoort ongeveer 30 procent lager dan een 5,5 en een zelfde percentage een 6,5 of hoger. Hoe gering de verschillen binnen een schooltype in werkelijkheid ook zijn, er zal op deze manier altijd blijken dat er verschillen tussen scholen zijn.

Voor het belang van openbare prestatiegegevens voor het schoolkeuzeprocess, het afleggen van rekenschap door scholen (*accountability*) en schoolverbetering zien Bosker en Scheerens allerlei haken en ogen.ⁱⁱⁱ De vraag is hoe belangrijk de overheid en ouders die prestatiegegevens vinden. Ook moet worden afgevraagd in hoeverre het schooleffectiviteitsonderzoek duidelijke aanknopingspunten biedt voor schoolverbetering. De empirische ondersteuning voor effecten van docenten en schoolkenmerken is immers beperkt. Voor schooldiagnose of zelfevaluatie is bovendien meer gedetailleerde informatie nodig. Bosker en Scheerens (1999) relativeren daarom de betekenis van het Trouw-onderzoek of de kwaliteitskaart. Ondanks deze relativeringen beschouwen zij openbare prestatiesgegevens van scho-

len als een nuttige praktijk die bijdraagt aan de rationaliteit van keuzegedrag, de besturingsprocessen op scholen en het overheidsbeleid. De aandacht voor prestatie-indicatoren is volgens hen een belangrijke prikkel voor scholen om meer opbrengstgericht te gaan werken. Een dergelijke wijziging in de schoolcultuur, zo verwachten Bosker en Scheerens, leidt tot prestatieverbetering.

Kwaliteitskaart en Trouw-onderzoek

De kwaliteitskaart van de onderwijsinspectie vermeldt vanaf het schooljaar 1999-2000 niet alleen het uitstroomniveau (eindexamencijfers), maar ook het instroomniveau (advies voor voortgezet onderwijs). Door zowel het instroom- als uitstroomniveau weer te geven, is volgens de inspectie te zien wat leerlingen op een school leren en of een school meer of minder dan gemiddeld bijbrengt. De speciale omstandigheden waaronder een school werkt, hoeven daarom niet meer te worden meegenomen. Er wordt dus niet meer gecontroleerd voor het percentage allochtone leerlingen of de urbanisatiegraad. Het ministerie volgt hiermee ten dele de kritiek van de Onderwijsraad (1998) op. Het advies voor voortgezet onderwijs is namelijk een minder vage en onduidelijke maat dan het percentage allochtone leerlingen of de urbanisatiegraad. Toch is het advies voor voortgezet onderwijs niet objectief. Allochtone leerlingen en leerlingen uit hogere sociale milieus krijgen gemiddeld genomen hogere adviezen dan hun score op de Cito-toets aangeeft (Dronkers, 1997a). Ook zijn er verschillen in advies naar regio en denominatie (Dronkers, Van Erp, Robijns & Roeleveld, 1998). Om het rendement van een school te meten blijft een aanvangstoets daarom het ideaal. Pas als het instroomniveau op die manier wordt bepaald en er multiniveau gegevens beschikbaar zijn, kan worden nagegaan wat leerlingen op een school leren en wat de toegevoegde waarde van een school is.^{iv}

Van het Trouw-onderzoek zijn inmiddels twee nieuwe versies verschenen (Agerbeek, 1998; 1999b). De berekening van de beoordeling in de tweede editie van het Trouw-onderzoek leek sterk op het eerdere rapportcijfer. In de tweede editie liepen de beoordelingen niet van één tot en met tien, maar werden de mogelijke scores uitgedrukt in plussen en minnen. De beide beoordelingen zijn in elkaar om te rekenen: een zes als rapportcijfer is bijvoorbeeld een nul als beoordeling en een vier of lager als rapportcijfer komt overeen met een dubbele min. Voor de tweede editie van het Trouw-onderzoek heeft het dagblad haar analyses gebaseerd op betere gegevens van scholen. Trouw vermeldt dat scholen de inspectieformulieren nu nauwkeuriger en op een meer vergelijkbare wijze hebben ingevuld. Zo worden de uitkomsten van leerlingen die buiten het reguliere onderwijs vallen, zoals leerlingen in internationale schakelklassen, niet langer

meegerekend. Voor de cijfers geldt bovendien dat ze (op het percentage allochtone leerlingen na) beschikbaar zijn van de afzonderlijke locaties en niet alleen van de school.

In de versie van 1999 heeft Trouw de gegevens van de nieuwe kwaliteitskaarten gebruikt. Met deze gegevens heeft het dagblad wederom een beoordeling met plussen en minnen opgesteld. Deze beoordeling is gebaseerd op de gemiddelde inspectiebeoordeling van drie elementen, te weten het aantal leerlingen dat onvertraagd in de derde klas belandt, het aantal leerlingen dat zonder in de bovenbouw zitten te blijven hun diploma haalt en het gemiddelde eindexamencijfer. Daarbij is rekening gehouden met het advies van leerlingen voor voortgezet onderwijs (en voor het rendement in de bovenbouw ook met het percentage allochtone leerlingen en het percentage leerlingen met een tegemoetkoming in de studiekosten).

Voor de analyses in de twee volgende versies van het Trouw-onderzoek zijn nog steeds alleen maar gegevens op geaggregeerd niveau beschikbaar. De belangrijkste bezwaren uit hoofdstuk 2 zijn dan ook nog steeds van kracht. Van een deugdelijke opzet voor de publieke verantwoording van scholen kan dan ook nog niet worden gesproken.

Menig een weet wel ongeveer dat voor verschillen in prestaties vooral kenmerken van leerlingen en gezinnen van belang zijn. Of veel mensen ook ervan uit gaan dat de invloed van scholen klein is, is de vraag. In de werkelijkheid verschillen scholen immers op talrijke punten van elkaar, bijvoorbeeld in de grootte van een locatie, de aard van de brugperiode en de inrichting van het mentoraat. Wanneer ouders voor de keuze van een school voor voortgezet onderwijs staan en de keuze hebben uit een aantal scholen, dan is de kans groot dat die scholen zich van elkaar onderscheiden. Dat onderscheid kan in de dagelijkse praktijk merkbaar zijn. Voor het eindresultaat, de prestaties van leerlingen of het rendement van de school, maakt het echter tamelijk weinig uit. Ouders hebben iets te kiezen, maar in dat keuzeproces gaat het er meer om dat de ene school *anders* is dan de andere en er minder om dat sommige scholen *beter* zijn.

De belangrijkste reden hiervoor is waarschijnlijk dat alle scholen hun programma richten op het niveau van het centraal schriftelijk eindexamen. Op die eindexamens bereiden scholen vwo-leerlingen zes jaar voor. Scholen kunnen niet variëren in de duur van dit traject. De enige variatie die mogelijk is, is dat op sommige vwo-scholen leerlingen vaker een extra examenvak nemen of dat op sommige vbo- en mavo-scholen meer leerlingen examen op het hoogste niveau afleggen. Verschillen tussen scholen van een zelfde schooltype maken dus weinig

uit voor de prestaties van leerlingen. Daarentegen zijn verschillen tussen schooltypen (hoofdstuk 4) wel van belang voor de prestaties van leerlingen.

Dit onderzoek werpt ook licht op de vraag naar ongelijkheid in het onderwijs. Hierbij blijkt dat deze vraag een genuanceerd antwoord verlangt. Wanneer wordt gekeken naar *verschillen in prestaties* valt op dat er grote verschillen zijn tussen leerlingen. In leerjaar 3 van het voortgezet onderwijs blijken allochtone leerlingen op Nederlands en wiskunde een achterstand te hebben op autochtone leerlingen. Behalve naar etniciteit zijn er ook duidelijke verschillen naar sociaal-economische status. Jongeren uit lagere sociale milieus scoren lager op de toetsen op beide vakken dan leerlingen uit hogere sociale milieus. Verder blijkt er ook sprake van sekse-ongelijkheid. Jongens presteren beter op wiskunde dan meisjes en, omgekeerd, scoren meisjes hoger op Nederlands dan jongens. Aan het einde van de leerplichtige leeftijd zijn er dus duidelijke verschillen op deze structurele kenmerken.

Wanneer echter naar *verschillen in vooruitgang* wordt gekeken, dan blijkt dat er geen invloed is van de kenmerken etniciteit en sociaal-economische status. Het enige structurele kenmerk dat een effect heeft op leervorderingen is de sekse. Meisjes blijken meer vooruit te gaan op Nederlands dan jongens. In leerjaar 3 van het voortgezet onderwijs blijken er aanzienlijke verschillen te zijn in prestaties, maar deze verschillen komen niet of nauwelijks tot stand door de vooruitgang die leerlingen in de onderbouw boeken. Ongelijkheid naar etniciteit en sociaal-economische status komt tot stand voordat leerlingen aan het voortgezet onderwijs beginnen en wordt niet vergroot tussen leerjaar één en drie, de periode van de basisvorming.^v In lijn met Dronkers (1997a) kan worden gesteld dat de verschillen die bij aanvang aanwezig zijn niet verder toenemen. Achterstandsbeleid voor leerlingen van een andere etnische afkomst of voor leerlingen uit lagere sociale milieus zal daarom gemiddeld genomen geen zin hebben in deze periode. Achterstandsbestrijding zou zich vooral op het leefpatroon van jongeren moeten richten. Van dergelijk beleid moeten geen grote wonderen worden verwacht. Culturele kenmerken, zoals het leefpatroon, hebben weliswaar een invloed op verschillen in relatieve posities, maar deze invloed is klein.

6.4 Naar vervolgonderzoek

In toekomstig onderzoek naar verschillen in prestaties acht ik een grotere *theoretische inbedding* wenselijk. In dit onderzoek heeft theorievorming niet centraal gestaan. Het hanteren van de structuur van concentrische cirkels en het vooraf op-

stellen van hypothesen is mijns inziens al een behoorlijke verbetering op exploratief onderzoek, waarbij allerlei verbanden achteraf worden verklaard.

Hoe moeilijk het is om voor onderzoek met een algemene vraagstelling een theorie op te zetten, blijkt uit het werk van Coleman. Uit een vergelijking van zijn publicaties over rationele-keuze theorie (Coleman, 1990) en onderwijsonderzoek (Coleman e.a., 1966; Coleman, Hoffer & Kilgore, 1982; Coleman & Hoffer, 1987) valt op dat zijn onderwijsonderzoek nauwelijks voortbouwt op zijn rationele-keuzetheorie. In de rapportages van Coleman over ongelijkheid in onderwijskansen of verschillen tussen openbaar en bijzonder onderwijs voert de empirie dan ook de boventoon over de theorie.

Voor meer specifieke vragen is een theoretische sturing eerder mogelijk. Hierbij kan de structuur van concentrische cirkels gebruikt worden als het algemene kader waarbinnen een specifieke vraag valt. Een voorbeeld hiervan is een onderzoek dat ik samen met Dijkstra heb uitgevoerd (Dijkstra & Veenstra, 1999; Veenstra & Dijkstra, 1998). In dat onderzoek ging primair de belangstelling uit naar de invloed van religieuze kenmerken van leerlingen, ouders en scholen op prestaties Nederlands, en fungeerden de kenmerken uit de structuur van concentrische cirkels als controlevariabelen.

In een meer theoretische benadering zou het uitgangspunt kunnen zijn dat het er voor prestaties van leerlingen om gaat waartoe ze in staat zijn (intelligentie, eerdere prestaties), wat ze er zelf aan willen doen (huiswerk maken, leefpatroon) en hoe ze worden beïnvloed in hun preferenties en intenties door hun zelfbeeld of door ouders, klasgenoten, docenten en de regels op school. De afwegingen die leerlingen maken en de beslissingen die ze nemen staan centraal, maar de structuur waarin ze zich bevinden kan beperkingen opleveren. Een dergelijke benadering zou passen binnen de *rationele-keuzetheorie*.

Rationele-keuzetheorie is gebruikt voor het verklaren waarom de schoolloopbanen van leerlingen verschillend verlopen na kritieke fasen. In een onderzoek naar de verschillende keuzemomenten in de schoolloopbaan van leerlingen heeft Boudon (1974) naar voren gebracht dat ouders bij elk van die keuzemomenten een afweging maken tussen kosten en baten. Deze afweging wordt volgens Boudon beïnvloed door het sociaal milieu van de ouders. Gambetta (1987) heeft een soortgelijke theorie gebruikt voor het verklaren van verschillen in voortijdig schoolverlaten. In de Nederlandse onderwijssociologie is vanuit dit gezichtspunt onder anderen onderzoek gedaan door De Vos (1986), die een model ontwikkelde om de invloed van de samenstelling van een schoolklas op leerprestaties te kunnen onderzoeken, Van Liere (1990), die onderzocht wat het advies voor voortgezet onderwijs is voor meer en minder regelcompetente leerlingen, en Maas (1992), die in groep vier van de basisschool de verwijzing van leerlingen naar het speciaal

onderwijs bestudeerde. Voor het verklaren van verschillen in prestaties is een dergelijke theorie voor zover bekend niet gehanteerd. Voor een algemene vraag naar verschillen in leerprestaties en leervorderingen tussen leerlingen zal het echter niet gemakkelijk zijn om een specifieke theorie op te stellen.

Behalve aan verdieping kan er in toekomstig onderzoek ook worden gedacht aan verlenging. Dit onderzoek richt zich op effecten op verschillen in prestaties over een periode van twee jaar. Het zou echter goed zijn om na te gaan wat de effecten op *langere termijn* zijn. In dit opzicht is het gunstig dat het CBS van de leerlingen uit VOCL'93 de loopbaan door het voltijdsonderwijs jaarlijks bijhoudt. Met de in hoofdstuk 2 besproken leerjarenladder als afhankelijke variabele kunnen er dus analyses worden verricht. Hierbij kan met de structuur van concentrische cirkels worden nagegaan wat de positie van leerlingen op de leerjarenladder na verloop van tijd is.^{vi}

Op dit onderzoek kan ook worden voortgebouwd door verbreding. Er kan namelijk naar meer dan invloeden op cognitieve uitkomsten worden gekeken. Nederlands en wiskunde mogen dan twee van de kernvakken van het voortgezet onderwijs zijn, toch zijn de prestaties die leerlingen daar op leveren niet de enige graadmeter voor het succes van leerlingen, docenten of scholen. Er zou bijvoorbeeld ook kunnen worden gekeken naar *andere uitkomsten*, zoals het probleemoplossend vermogen van leerlingen of de mate waarin ze beschikken over sociale en communicatieve vaardigheden (Trier & Peschar, 1995).^{vii} Voor de toetsen Nederlands en wiskunde geldt dat ze tamelijk traditioneel zijn. In de toetsen wordt bijvoorbeeld geen aandacht besteed aan leren leren, het gebruik van moderne media, of de mate waarin leerlingen kunnen samenwerken.

De verwachting is dat factoren die op leerlingniveau cognitieve uitkomsten beïnvloeden doorgaans ook van invloed zijn op non-cognitieve uitkomsten. Uit onderzoek van Grisay (1996) in Frankrijk blijkt dat non-cognitieve uitkomsten sterk worden beïnvloed door leerling- en gezinskenmerken. Verder blijkt de relatieve tussenschoolse variantie in non-cognitieve uitkomsten kleiner dan in cognitieve uitkomsten. Wat de vaardigheden en houdingen van leerlingen betreft zijn er voornamelijk verschillen binnen en nauwelijks verschillen tussen scholen. Hoewel eerdere bevindingen (Creemers, 1991) erop wijzen dat er geen sterke relatie bestaat tussen affectieve en cognitieve onderwijsuitkomsten, blijkt uit het onderzoek van Grisay dat schoolkenmerken die van invloed zijn op cognitieve opbrengsten in grote lijn ook van invloed zijn op affectieve opbrengsten.

Een andere mogelijkheid is om naar keuze- en selectiemomenten in het voortgezet onderwijs te kijken (Dekkers, Bosker & Meijnen, 1998). In dat kader kan worden gekeken naar de keuze van leerlingen voor een schooltype na de brugperiodes, de keuze van een richting voor leerlingen in het beroepsonderwijs of van

een vakkenpakket voor leerlingen in het algemeen vormend onderwijs (Dekkers, 1996), en naar de selectie die optreedt door voortijdig schoolverlaten (Bosker & Hofman, 1994; Dekkers & De Wit, 1995; Gambetta, 1987). De vraag in welk schooltype leerlingen na de brugklas terechtkomen is bijvoorbeeld uiterst relevant. In hoofdstuk 4 bleek dat 37 procent van de leerlingen in een dakpansgewijze en 10 procent in een heterogene brugklas zitten. Voor deze leerlingen kan met logistische regressie worden nagegaan waarom sommige leerlingen doorstromen naar het hogere schooltype en andere naar het lagere schooltype. Voor onderzoek naar dergelijke keuzemomenten is de structuur van concentrische cirkels waarschijnlijk ook toepasbaar.

Ook verruiming behoort tot de mogelijkheden voor vervolgonderzoek. De structuur van concentrische cirkels zou ook in onderzoek met *andere datasets* gebruikt kunnen worden. Hierbij kan worden gedacht aan onderzoek in andere typen van onderwijs en aan onderzoek in andere landen.^{viii} Voor een validatie van deze structuur is het noodzakelijk om nieuwe gegevens te gebruiken (Freedman, 1991, p. 307). "Generally, replication and prediction of new results provide a harsher and more useful validation regime than statistical testing of many models on one data set. Fewer assumptions are needed, there is less chance of artifact, more kinds of variation can be explored, and alternative explanations can be ruled out. Indeed, taken to the extreme, developing a model by specification tests just comes back to curve fitting."

Wat de verdieping, verlenging, verbreding of verruiming ook mag inhouden, er zal rekening moeten worden gehouden met de technische vooruitgang die ik in paragraaf 6.2 heb beschreven. In onderwijsonderzoek zal daarom moeten worden gekozen voor multiniveau analyse.^{ix} Wanneer er verschillende afhankelijke variabelen beschikbaar zijn, kan het beste van de voordelen van een multivariaat model gebruik worden gemaakt. Om over effecten op hogere niveaus uitspraken te doen dient verder zo veel mogelijk rekening te worden gehouden met effecten op lagere niveaus. In onderzoek naar effectieve docenten en scholen, moet dus voor zowel structurele als culturele leerling- en gezinskenmerken worden gecontroleerd. Wanneer dergelijke kenmerken in samenhang worden bekeken, verdient het aanbeveling om bijvoorbeeld de structuur van concentrische cirkels te hanteren. Ook geldt dat als het om onderzoek in het voortgezet onderwijs gaat, het verstandig is om analyses voor alle schooltypen tegelijkertijd uit te voeren. Voor sociaal-wetenschappelijk onderzoek geldt tot slot dat er niet alleen naar invloeden op een toestand op een bepaald moment (leerprestaties), maar ook naar invloeden op verandering tussen twee momenten (leervorderingen) moet worden gekeken.

Eindnoten

ⁱ In Groot-Brittannië hebben meisjes sinds kort niet alleen in taal maar ook in wiskunde een voorsprong op jongens (Sammons, 1995). Uit dit onderzoek blijkt dat in Nederland meisjes nog geen algehele voorsprong hebben genomen.

ⁱⁱ Vergelijk het belang dat Harris (1998) aan vriendengroepen toeschrijft voor de ontwikkeling van de persoonlijkheid.

ⁱⁱⁱ Uit onderzoek van Trouw blijkt echter dat de rapportcijfers van Trouw een rol spelen bij de schoolkeuze (Agerbeek, 1999c). Voor het schooljaar 1998-'99 was de aanmelding op alle scholen gemiddeld genomen 9 procent hoger dan het jaar daarvoor. Op scholen met een rapportcijfer hoger dan een zeven steeg de aanmelding echter met 33 procent.

^{iv} Over openbare ranglijsten van scholen merkt Waslander (1999) op dat er in deze discussie niet alleen sprake is van een informatieprobleem (welke gegevens zijn nodig om de kwaliteit van scholen weer te geven?) maar ook van een kennisprobleem. Om een 'wijze' schoolkeuze te maken, moeten ouders ook beschikken over kennis van het onderwijssysteem, het functioneren van scholen, de aanmeldingsprocedures voor scholen, et cetera. Dit probleem, zo stelt zij, is niet op te lossen met meer informatie, maar vereist een meer omvattende aanpak van sociale ongelijkheid.

^v Van der Werf, Lubbers en Kuyper (1999) stellen dat sinds de invoering van de basisvorming de mate van ongelijkheid naar sociaal-economische achtergrond en etniciteit niet is afgenomen. Zij baseren zich hierbij op analyses met de covariantie-benadering. Met die benadering kunnen zij niet tot deze slotsom komen. De basisvorming moet worden beoordeeld op de vooruitgang van leerlingen tussen leerjaar één en drie en niet op verschillen die mogelijk in de periode voor het voortgezet onderwijs zijn ontstaan.

^{vi} Door cohorten lange tijd te volgen ontstaan mogelijkheden voor prachtige analyses. Befaamd is in dit opzicht het Malmö-cohort, waarin 1.544 kinderen die in 1938 tien jaar waren vijftig jaar lang zijn gevolgd. Met dit cohort hebben Husén (1969), Fägerlind (1975) en Tuijnman (1989) onder meer laten zien dat extra gevolgd onderwijs (her-, bij- of nascholing) op lange termijn invloed heeft op het bereikte beroepsniveau, het inkomen en de ervaren gevoelens van tevredenheid. Hierbij blijkt dat dergelijke effecten zich op 35-jarige leeftijd nog nauwelijks voordoen, maar op 56-jarige leeftijd bijzonder sterk zijn. Het effect van extra kwalificaties is onafhankelijk van andere kenmerken, zoals het sociaal milieu, de intelligentie en het initiële opleidingsniveau van leerlingen.

Een dergelijk cohort is uniek. In Nederland zijn er niet zulke langlopende cohorten, maar er zijn wel enkele cohorten die op een later tijdstip opnieuw zijn aangewend. Peschar (1975) heeft begin jaren zeventig opnieuw informatie verzameld bij personen die in het schooljaar 1958-'59 in de hoogste klas van het lager onderwijs zaten. Dronkers (1998b) heeft gebruik gemaakt van het gereactiveerde cohort van Matthijssen en Sonnemans (1958).

Voor onderzoek naar lange-termijneffecten zou het gunstig zijn, wanneer in Nederland de verschillende cohorten meer op elkaar aansluiten. In het basisonderwijs is er het PRIMA-cohort; in het voortgezet onderwijs is VOCL'93 het meest recente cohort en in het hoger onderwijs is er Verder Studeren. Tussen deze cohorten is echter geen overlap. Het zou naar mijn mening de moeite waard zijn, wanneer het Ministerie van Onderwijs, Cultuur en Wetenschappen en het Ministerie van Sociale Zaken en Werkgelegenheid zich hard zouden maken voor een lang lopend onderzoek naar verschillen in school- en beroepsloopbanen. Ook is het wenselijk om een data-infrastructuur te ontwikkelen waarbij verschillende bestanden aan elkaar kunnen worden gekoppeld via een landelijke basisregistratie (Allen, Heijke & Van der Velden, 1999). Door

koppeling van gegevens uit verschillende bronnen wordt tijd en geld bespaard, onnodige overlap vermeden, het onderzoeksveld ontlast, een ver gaande coördinatie tussen onderzoeken bevordert, en de informatiewaarde van verzamelde gegevens aanzienlijk verhoogd.

^{vii} In het verleden zijn talrijke non-cognitieve uitkomsten als afhankelijke variabele gehanteerd, zoals de mate van delinquentie (Rutter e.a., 1979), sociale vorming (Van Marwijk Kooy-von Baumhauer, 1984) en affectieve en sociale ontwikkeling (Bryk, Lee & Holland, 1993).

^{viii} Hierbij zou het ook zeker voorkeur verdienen om onderzoeksinstrumenten meer te standaardiseren. Replicatie wordt in onderwijsonderzoek vaak bemoeilijkt doordat er te veel verschillende instrumenten worden gebruikt.

^{ix} In toekomstig onderzoek moet ook zeker nog meer naar interacties tussen leerling-, klas- en schoolkenmerken worden gekeken. In dit onderzoek bleek dat sommige schoolkenmerken bijvoorbeeld wel van invloed zijn op leerlingen in het beroepsonderwijs, maar niet op leerlingen in het algemeen vormend onderwijs. Op soortgelijke wijze kan er ook worden gekeken naar interacties van sekse, etniciteit, sociaal-economische status of culturele leerling- en gezinskenmerken met kenmerken op klas- of schoolniveau.

Summary

Pupils – classes - schools

Using data on around 7,000 pupils from 450 classes in 150 schools (information from the cohort study VOCL'93), I have studied the extent to which differences in achievements and progress in Dutch and mathematics may be explained through characteristics of pupils, families, teachers and schools. I have focussed on the achievements and progress of pupils in lower secondary education in the Netherlands. The central questions are these:

- 1 To what extent do the estimates of the net achievements of schools differ between an analysis based on aggregated data and an analysis using a multilevel statistical model?
- 2a To what extent and why do characteristics at an *individual level* influence the *achievements* of pupils in the third year of secondary education?
- 2b To what extent and why do characteristics at *class and school level* influence the *achievements* of pupils in the third year of secondary education?
- 3a To what extent and why do characteristics at an *individual level* influence the *progress* (differences in the relative learning gain) of pupils between the first and the third year of secondary education?
- 3b To what extent and why do characteristics at *class and school level* influence the *progress* (differences in the relative learning gain) of pupils between the first and the third year of secondary education?

To understand the structure of Dutch secondary education, some general remarks might be helpful. Children start their education at the age of four and leave primary education about eight years later. Then they will enter secondary education, which is divided into five tracks: individual lower vocational education (IVBO; the duration is four years), lower vocational education (VBO; four years), lower general secondary education (MAVO; four years), intermediate general secondary education (HAVO; five years), and pre-academic education (VWO; six years). Each school track for secondary education has its own subsequent track. Pupils from pre-academic education are allowed to go to the university. However, one quarter of the pupils chooses higher professional training (HBO colleges) instead of university. A HAVO-certificate is also sufficient for admission to HBO colleges. The subsequent track for VBO or MAVO graduates is a junior professional training (MBO colleges).

A nested structure (question 1)

Multilevel analysis has various advantages over analyses at one level, such as a regression analysis with aggregated or disaggregated data. By taking into account the hierarchy of data (pupils within classes within schools), the results are more accurate and may be interpreted easier than the results from an analysis at one level. Another advantage of multilevel analysis is that differential effects may be studied. For example, the effect of ethnicity on achievements is not the same in each class. In some classes this effect is stronger and in others it is weaker. This means that one class may on average be more advantageous for an ethnic minority pupil and another class may be relatively less advantageous. A third advantage of multilevel analysis is that it can distinguish between regression within and between groups.

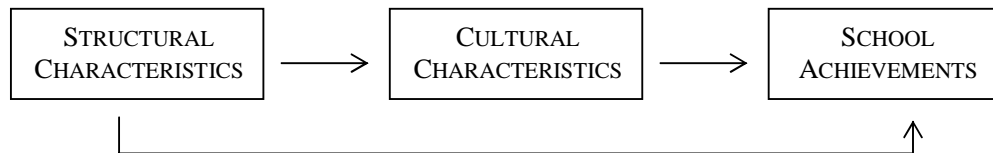
Recently, achievement data on schools have been made publicly available. In 1997 the newspaper Trouw published such information for the first time. A year later the government published similar data. These forms of information on schools raise some questions. For one thing, only data at an aggregated level are available. In a replication of the Trouw-study of school achievements I illustrated that the results of analyses at one and two levels differed significantly from each other. In more than half of the schools, the grades assigned differ one or more points (on a scale ranging from four to eight) between an analysis with aggregated data and a multilevel analysis. If the researcher does not take the nested structure into account, the results should at least be interpreted with a great deal of care. Measures of success from the one level analysis, such as the marks in the Trouw-study, will therefore not give a reliable picture of the difference in achievements between schools.

The results also differ significantly between analyses of gross and net achievements. An indicator of the gross achievements of a school is, for instance, the average level of competence in Dutch and mathematics. In order to give a fair picture of the way in which 'schools matter', it is necessary to take into account the circumstances of a particular school. The net achievements of a school, in which these circumstances are controlled for, is the average level of competence of pupils in these subjects, after taking into account differences in the composition of pupil populations. Using data at one level, as was done in the Trouw-study, it is not possible to take composition differences into account in a fully satisfactory way. However, when data are collected at both the individual and the school level, the net achievements of a school may be determined in a multilevel analysis.

Pupils and families differ

Many researchers have investigated the influence of pupil and family characteristics on achievements. In many of these studies researchers do not wish to restrict themselves to describing educational inequalities. More and more researchers seek to find out *why* there are differences between pupils from lower and higher social classes, between boys and girls, between pupils of ethnic minorities and native Dutch pupils, and between children from single-parent families and dual-parent families. However, as yet the cause of differences in the achievements of pupils can only be explained in part. One of the reasons for this might be that most researchers investigate the influence of just a few characteristics on achievements. Therefore, researchers are faced with the challenge of investigating the many characteristics of families and pupils in conjunction.

When explaining differences in pupils' achievements, the starting point should be the important position of pupils and their families. Differences in achievements may be explained in particular by characteristics at an individual level. On this level a distinction may be made between *structural* and *cultural* characteristics. Cultural characteristics, such as the life style of pupils or the parenting style in the family, may partially intermediate the influence of the structural characteristics, such as the socioeconomic status and the ethnicity (see the figure below).

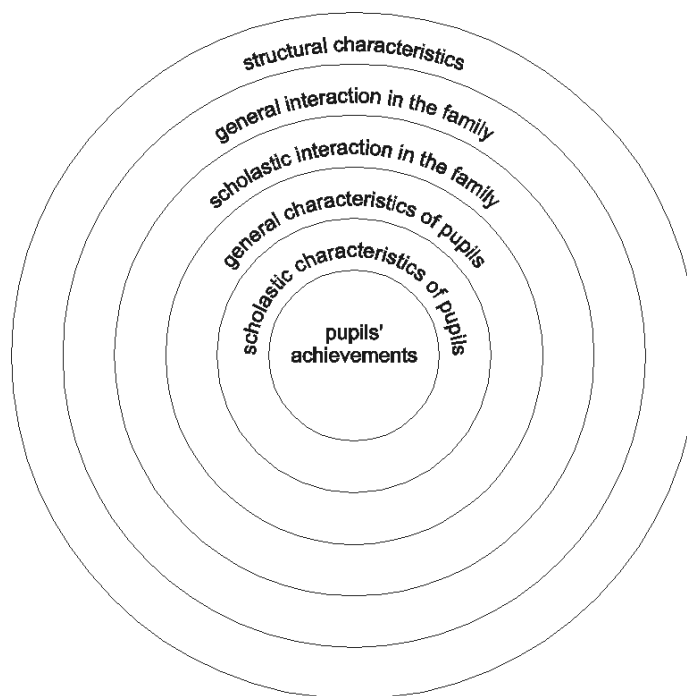


In the figure above it is clear that the influence of structural characteristics on the achievements of pupils may be either direct or indirect, via cultural characteristics. This model can be used to determine whether the influence of structural characteristics can be explained by cultural characteristics. Cultural characteristics may be subdivided into pupils and family characteristics, and into scholastic characteristics and more general characteristics. A combination of these distinctions results in four categories of cultural characteristics: general family characteristics, scholastic family characteristics, general pupil characteristics and scholastic pupil characteristics (see diagram below).

	FAMILY	PUPIL
GENERAL	4	2
SCHOLASTIC	3	1

These four categories may be arranged according to the extent to which they are directly related to the pupils' achievements. The first assumption in this arrangement is that characteristics of pupils are more related to their achievements than family characteristics. The second assumption is that scholastic characteristics in both pupil and family characteristics are more directly related to the achievements than are the general characteristics. This results in an arrangement of characteristics in which scholastic characteristics of pupils are most intimately related to the pupils' achievements and general family characteristics are least directly related to these achievements.

The structural and cultural characteristics are the independent variables which explain the differences in achievements. Apart from the independent variables, one can also distinguish dependent variables. In this study the dependent variables are the grades for a Dutch test (text comprehension) and a mathematics test.



I distinguish a total of six categories of characteristics at an individual level: one category with the dependent variables, one category with structural

characteristics, and four categories with cultural characteristics. These six categories form a *structure of concentric circles* (see figure above). The structural characteristics form the outermost circle (the fifth) and the dependent variables, the pupils' achievements, form the core. The cultural characteristics have been placed as intermediate variables between the structural characteristics and the dependent variables. The general interaction in the family, which consists of the parenting style, forms the fourth circle. The scholastic interaction in the family, for instance homework monitoring, comprises the third circle. The second circle consists of the pupils' general characteristics, such as their activities in their spare time. The scholastic characteristics of pupils form the first circle. These comprise, for instance, a pupil's earlier achievements and intelligence.

Using this model I aim to arrange the various influences and to offer more insight into the (causal) relations between the characteristics. It would also seem possible to explain the effects of characteristics such as the socioeconomic status or the ethnicity on school achievements by means of intermediate variables. An attempt will be made to explain the effects of characteristics from the outer circles by means of characteristics from the inner circles. The influence of characteristics in one circle will be regarded as a whole. This means that no investigation will be conducted on the possible arrangement of connections.

The individual level: achievements (question 2a)

The achievement levels of pupils in Dutch and mathematics are determined by both structural and cultural characteristics. Even if the prior achievements of pupils are taken into account, it is important to know the *education of the parents* and whether they belong to *ethnic minorities* or to *the native-born majority*. When the effect of these characteristics is studied in conjunction, there is a strong effect. A native-born pupil whose parents have received a higher professional education, will score one quarter of a standard deviation higher than a pupil of an ethnic minority whose parents have received only a basic education. Comparing this to the difference in achievements between pupils in different school types, for example a pupil in lower and a pupil in intermediate general secondary education, this is around one half of a standard deviation. The difference between pupils of ethnic minorities and native-born pupils does not disappear if the characteristics of the socioeconomic status are taken into account. Ethnicity has therefore an extra effect on achievements, not only in Dutch, but also in mathematics.

If the other characteristics are kept constant, *gender* also has a strong effect on the achievements of pupils. Girls have a better average score on Dutch and boys have a better average score on mathematics. Growing up in a *dual-parent family* positively influences the achievements in mathematics.

In addition to the effects of structural characteristics there are also effects of cultural characteristics. *Homework monitoring* and an *authoritative parenting style* are forms of interaction between parents and children which have a rather strong connection to the pupils' achievements. Pupils who are not urged on and whose homework is not monitored by their parents, and pupils who make 50 percent of their decisions after consulting their parents (meaning that the authoritative parenting style is regularly used), have scores on Dutch which are 0.33 of a standard deviation higher for Dutch and 0.18 of a standard deviation higher for mathematics than the scores of pupils who are subjected to homework monitoring and who are not raised in an authoritative manner.

One of the general characteristics of pupils with a strong positive effect is the pupils' self-concept. This effect is as strong as the combined effect of the education of both parents. The *life style* also has a strong effect on the achievement of pupils. Pupils who spend half their spare time going out, working or hanging out with friends, and who do not spend any time on reading or hobbies, have scores a quarter of a standard deviation lower on Dutch, and one fifth of a standard deviation lower on mathematics than pupils who spend 20 percent of their time on reading or on hobbies and who do not spend time on the three other aspects.

Naturally, the strongest effects on the achievements stem from the effects of *earlier achievements*. For instance, when a pupil has grades for all three tests (reading, information processing, and mathematics) which are one standard deviation above average, this pupil's achievements are at HAVO-level two years later. At two standard deviations above average, the pupil's grades are at VWO-level in the third year. A difference of one standard deviation is roughly equivalent to a difference in school type. Thus it is clear that the pupils' prior achievements are a good predictor of their later achievements. With regard to the other characteristics in circle one, the number of days that a pupil is given homework to do is of considerable influence. When a pupil does homework for only two days instead of five, the achievement in tests drops by one tenth of a standard deviation.

When the various circles are compared to one another, it appears that the achievements of pupils may be explained better through pupil characteristics than through family characteristics. In the survey of the literature many more pupil characteristics emerge than family interaction. It also appears that the explained variance increases much more by including the pupil characteristics (circles 1 and 2) than by including the family characteristics (circles 3 and 4). As far as the pupil characteristics are concerned, it turns out that the pupils' achievements may be better explained through school related characteristics than through general characteristics. This conclusion cannot be drawn for family characteristics. Furthermore, cultural characteristics are more important than structural

characteristics for the explanation of differences in achievements. The structural characteristics account for less than 10 percent of the variance. The cultural characteristics explain four to five times as much variance.

In addition to these pupil and family characteristics, the type of school is also influential. Not only does it provide an added effect, but it also mediates other effects, such as the effect of the education level of the parents or the number of days the pupils do homework.

The individual level: progress (question 3a)

In the previous paragraphs the central question was the extent to which the achievements of pupils at a certain moment in time could be explained by various characteristics of pupils and families. In order to determine the net achievements of pupils, the achievements of the pupils at an earlier time are taken into account. The test scores in year three are the dependent variables and the prior achievements in year one are included in the independent variables. This is a standard model in studies of school effectiveness and is referred to as the *covariance-analytical model*.

Using this model, pupils' achievements at a certain point in time can be explained. In this explanation the question is which pupils are *currently* behind or have a lead. Then the characteristics which influence the learning *achievements* are identified. This model is therefore used to study influences on learning achievements, not to study influences on learning progress. When a study is based on this model, the researcher cannot speak of effects on progress in achievements. In order to study the progress in achievements, the researcher should choose a *variance-analytical model of differences in relative learning gain*. The dependent variables in this model are for instance *the differences* between the standardized achievement scores in year three and year one.

The results of the covariance approach and the learning gain approach turn out to diverge strongly at the individual level. Structural characteristics (ethnicity, level of education of father and mother, or the number of parents) and part of the cultural characteristics (authoritative parenting style, spending time on reading and hobbies, opinion of class mates and non-verbal intelligence) do have an effect on the test scores for Dutch and mathematics, but have no effect on the differences in the relative position on these subjects. Conversely, the analysis of differences in the relative positions shows an effect of a cultural characteristic which accounts for the achievement motivation. That effect is absent in the covariance approach. Besides these pupil and family characteristics, the school type also has a weak influence on the progress pupils make.

The analyses show that structural characteristics hardly have any effect on the relative position on both subjects. Contrary to the common supposition, the disadvantage of pupils of ethnic minorities does not increase during secondary education. Various cultural characteristics do have an effect on the progress in Dutch and mathematics. The changes in achievements stem from pupil characteristics, such as their commitment to school and their life style. From a meritocratic point of view this result is favorable. Characteristics which are a given, such as the socioeconomic status and the ethnicity, do not influence differences in relative positions, but cultural characteristics do.

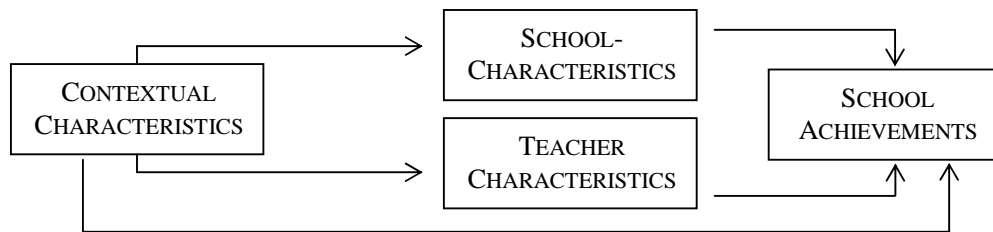
The influence of these changeable characteristics should not be overestimated. The cultural characteristics explain little of the variance in differences in relative positions. Therefore, individual characteristics hardly influence the learning gain. When entering secondary education, pupils take in a certain position as far as achievements are concerned. These positions hardly change during the first three years of secondary education (the correlation between the achievements in year one and three is 0.55 for Dutch and 0.74 for mathematics).

Teachers and schools differ

The effect of school characteristics is small compared to individual characteristics. Still, there are a number of reasons for stating that schools do matter. It appears that schools with good achievements in a particular year perform well the next year. Researchers remark regularly on the practical significance of school effects. Even small effects may be relevant, if only because the effects should be multiplied by the number of pupils benefiting from outstanding schools. Finally, in research on the effects of school characteristics the main issue is whether the *quality* of schooling makes a difference, not the question of whether the *quantity* of schooling is important. However, the quantity of schooling appears to be important for the pupils' achievements.

To what extent do characteristics of schools and teachers explain differences in achievements? Much research has been undertaken in this field. In particular, educational economists have studied the material school characteristics. The results of various studies are inconsistent and the effects of the material characteristics are weak at the most. The influence of immaterial school characteristics is the central issue in the *study of school effectiveness*. School effectiveness usually means the component of progress of pupils which cannot be ascribed to influences other than those of the school itself. This means that in school effectiveness studies the net achievements of schools is studied. The researchers look into the process characteristics of schools, such as the orderliness and safety of the school climate and a clear accent on the pupil's acquisition of the

basic skills. In the Netherlands the study of school effectiveness is receiving a lot of attention. Besides characteristics of effective schools, some researchers also study characteristics of effective teachers. This type of research is known as the study of *effective instruction* and it focuses on differences between classes. In this type of research characteristics such as the presence of the opportunity to learn and an effective learning time, structured education, high expectations about achievements, and stimulation are central issues. Finally, there are studies in which *contextual school characteristics* are central, such as research on the differences between public and private education.



There are three types of characteristics at the class and school level: teacher characteristics, school characteristics, and contextual characteristics. In various studies of effectiveness the importance of these types of characteristics has been pointed out. As the figure above shows, contextual effects may be explained by teacher and school characteristics, just as the effects of structural characteristics can be partly explained by cultural characteristics. In the study of the differences between public and private education the *black box* of the school may be said to be opened. It is not the label Catholic which explains the success of Catholic schools in the United States, but elements like the emphasis on basic skills or the school climate. This type of interpretive research aims to let teacher and school characteristics intermediate the influence of contextual characteristics as much as possible.

The class and school level: achievements and progress (questions 2b and 3b)

At the class and school level the similarities between the approaches of covariance and difference scores are stronger. The characteristics which influence the achievements in year three and which influence progress between year one and three are on the whole comparable.

In the Netherlands the gross variance in achievements at the school level turns out to be high. In this study the variance is 29 percent for Dutch and 38 percent for mathematics. Between classes and schools jointly these values are even higher: 51

and 65 percent. The variance between classes and schools is so high, because in Dutch secondary education pupils are ordered according to school type. In *Education at a glance* (OECD, 1998) it is stated that the Netherlands is an international frontrunner in interschool variance in secondary education. Other countries with high scores in this respect are Ireland, Germany, Switzerland, Austria and the Flemish part of Belgium. In contrast, in Korea and Japan the gross variance between classes and schools is less than 5 percent. Incidentally, this measure does not say anything about the inequality in achievements within a country. As it turns out, there is no clear relationship between the interschool variance (a relative measure) and the overall variance in achievements (an absolute measure).

The gross variance is not a measure of the difference in quality between schools. In order to compare schools with each other, the net variance should be computed. The net variance in achievements (in which value the pupil and family characteristics and the school type the pupil attends are taken into account) between schools is 12 percent for Dutch and 19 percent for mathematics. The net variance at the school level in the learning gain approach is 9 percent for Dutch and 12 percent for mathematics. In this approach too, only a small part of the variance in the gross variance is related to the systematic differences between schools. The composition of schools is not as strongly related to study progress as it is to study achievements. Furthermore, in both approaches there is more interschool variance for mathematics than for Dutch. This is caused by the fact that the relationship between what happens at school and mathematics achievement or progress is stronger than the relationship between what happens at school and the achievement or progress in text comprehension.

Of the *contextual characteristics*, the percentage of pupils of ethnic minorities and the size of the school location both influence the *achievements* of the pupils. For pupils in all school types the percentage pupils of ethnic minorities has a negative effect on the achievements in mathematics. An increase of one standard deviation in the percentage of pupils of ethnic minorities is associated with a decrease in the mark of 0.03 of a standard deviation. The achievements of pupils in Dutch on a HAVO or VWO school are also lower as the percentage of pupils of ethnic minorities rises. For each standard deviation difference in the percentage of pupils of ethnic minorities, the grades for the test change by 0.09 of a standard deviation.

For pupils in professional education the size of the school location accounts for a difference in the achievements in both subjects. When the number of pupils in a location rises from 300 to 600 or from 600 to 1200, this leads to a decrease in achievements of 0.07 of a standard deviation for both subjects. Since many pupils

undergo the consequences of such a school effect, it should be considered relevant. For pupils in MAVO, HAVO and VWO schools the size of the location does not influence the achievements.

In research outside the Netherlands, differences in achievements are regularly demonstrated between public and private schools. These differences occur in gross and net achievements. The current study shows that in the Netherlands differences occur in net achievements between denominations. Contrary to American studies, for instance, public schools do not perform worse than private schools. The differences occur mainly in the private sector. The arrears of pupils at interdenominational schools are conspicuous in this connection.

The *progress* of pupils is not influenced by the percentage of pupils of ethnic minorities, but it is associated with the size of the location and the denomination. However, the percentage of pupils of ethnic minorities is related to the achievements of pupils in year three of secondary education. This relationship is probably related mainly to the disadvantage of pupils at schools which had many pupils of ethnic minorities at the start of their secondary education, and not to the processes in the first three classes of secondary education. This is indicated by the fact that the disadvantage of pupils at 'black schools' does not increase between years one and three.

In the covariance approach the location size only has a negative effect on the achievements of pupils in professional education. In the learning gain approach a negative effect for all pupils is found. This effect indicates that the processes at a school differ according to the location size. There is also a denomination effect. Again, there are no significant differences between public, Protestant and Catholic schools. Interdenominational schools score negatively again. Pupils at this type of school incur arrears in mathematics.

Some of the *school characteristics* have significant effects. The stability of the team of teachers has a weak positive effect on the study achievements in both subjects. Each year that the team remains unchanged increases the achievements of pupils by one-twentieth of a standard deviation. This characteristic has no effect on study progress.

When schools substitute a canceled class by having another teacher take over or an obligatory homework class, the scores on Dutch of VBO or MAVO pupils are one-fifth to one-third higher than in cases where this class was canceled altogether. Substitution of lessons has an effect on the progress of VBO/MAVO pupils, but the effect of an obligatory homework class is not significant in the learning gain approach.

Pupils (except VWO pupils) in heterogeneous first classes have lower grades on Dutch in both the covariance and the learning gain approach than pupils in a homogenous first class or a so-called overlapping first class (a class in which pupils of at most two school types are brought together). When pupils of three or more school types were in the same class in the first year of secondary education, these pupils incur arrears for Dutch in the third year.

In both approaches the setup of the mentor system has the following characteristics: the more extensively this system has been set up, the lower the pupils' achievements. All forms score lower than a narrow, vertical mentor system, the reference group. Finally, it emerges that (I)VBO pupils perform better in a lower teacher-pupil ratio. Like pupils of IVBO, VBO-pupils benefit from a structure with not too many different teachers. In the learning gain approach this effect is only found for Dutch. In general these analyses showed it was hardly, if at all, possible to explain the influence of contextual characteristics by school characteristics.

Scientific relevance

The conclusion that pupil and family characteristics are decisive factors for differences in achievements accords with the research of Coleman et. al. (1966) and Jencks et. al. (1972). In thirty years time not much has changed in this respect. Technically there has been some progress, because of the development of *multilevel analysis*. This type of analysis has several advantages over one-level analyses of hierarchical structures.

It has not been usual in educational research to combine several achievement variables simultaneously in a *multivariate model*. In the current study, the analyses for Dutch and mathematics were conducted concurrently. This has the advantage that it facilitated a test of whether explanatory variables have the same effect on these two dependent variables. Moreover, the correlations between the dependent variables at various levels may be determined. Pupils who lack a certain value for one of the dependent variables may be included in the study, so incomplete data can be used more efficiently. A multivariate, multilevel model has important advantages over the models which were used a few decades ago. The development of these models enables better and more extensive research on differences in study achievements and study progress between pupils.

The development of multilevel analysis makes it possible to study the influence of pupils, families, teachers, and schools in *conjunction*. The current study shows the importance of verifying the characteristics at an individual level. In studies of effective teachers and schools the verification is often reduced. Researchers of effectiveness take structural characteristics and earlier

achievements of pupils into account, but they disregard cultural characteristics. However, in order to explain differences in achievements between pupils, these cultural characteristics are at least as important as structural characteristics. For an explanation of an increase or a decrease in relative position, structural characteristics are hardly important, if at all, whereas cultural characteristics do play a role. In order to make statements about effects at the class and school level the researcher should take both structural and cultural characteristics into account.

In a study with various characteristics at various levels, the danger exists that the research is directed entirely by empiricism. The researcher drags the data forward with a trailnet and after various analyses certain characteristics remain stuck in the net. These characteristics are then supposed to influence the dependent variable. In this study I have endeavored to avoid this trap by using a *structure of concentric circles*. Based on theoretical considerations, characteristics have been included in this structure. Using this structure analyses have been carried out per circle. The structure enables an organization of various influences and an indication of (causal) relations between characteristics. Effects, such as the mother's education or the ethnicity, can be explained for an important part by characteristics with a more direct relation to achievements. Around three-fifths of both effects can be explained by cultural characteristics. By adding the school type, the effects of the mother's and the father's education are explained entirely. Therefore, the structure of concentric circles seems particularly suited to *explaining* the differences in achievements at the individual level.

At the class and school level the analyses partly meet the expectations. At both levels there are a number of characteristics which were expected to be important for the differences in achievements in lower secondary education in the Netherlands. These expectations are in part confirmed empirically. For a number of characteristics, especially teacher characteristics, the measurement instruments from VOCL'93 are not suited. The current study can shed no light on the characteristics of effective teachers.

In the analyses IVBO, VBO, MAVO, HAVO and VWO pupils were studied in combination. It is technically feasible to carry out the analyses for *all school types at once*. This has the advantage that the number of research elements is not diminished unnecessarily. Moreover, by including interactions, effects may be studied which are specific to a certain school type. Therefore, a study of the characteristics of effective teachers and schools does not necessarily have to be carried out according to school type.

Finally, there are differences at the individual level between the results of the *covariance approach and the learning gain approach*. A covariance analysis does not provide an answer to the question of which pupils achieve a learning gain. An

increase or decrease in pupils' achievements can only be studied by an analysis of learning gain or relative positions. A comparison of both analytic methods reveals that they complement one another. The selection of one of these methods depends on the research question. When a researcher is interested in the question of which pupils have incurred a disadvantage or a lead at a certain point in time, the covariance analysis is called for. When a researcher wants to know which pupils incur a disadvantage or a lead in the course of their education, the right choice is an analysis of learning gain or relative positions. In contrast to a study of the characteristics of effective pupils and families, the choice of model in a study of the characteristics of effective schools is not very important.

Both technically and substantively, there has been progress in research on differences in achievements. The answer to the question of *differences in achievements* may be answered with more force and nuance. In order to explain differences in achievements, characteristics at an individual level are crucial. Of these characteristics the influence of cultural characteristics is much larger than that of structural characteristics. Of the cultural characteristics the influence of scholastic pupil characteristics is largest, followed by general pupil characteristics and the family characteristics.

To explain *differences in relative positions*, characteristics at an individual level are much less suited. With the exception of gender, structural characteristics do not influence differences in relative positions. Various characteristics do influence these differences, and the effects of pupil characteristics are stronger in this regard than those of family characteristics. There is no clear difference in the strength of effects between general and scholastic characteristics. As far as explained variance is concerned, it is striking that the characteristics at an individual level hardly explain any variance. It may be concluded that little is known about why certain pupils in lower secondary education incur a lead while others fall behind.

Social relevance

Newspapers publish achievement data on schools as lists. The government publishes a quality map with data about the exit level (examination results) and the entry level (recommended track of secondary education). However, these forms of publicly available information do not provide a sufficient basis for evaluating school effects. For one thing, only data at an aggregated level are available. Moreover, the initial level of pupils can only be taken into account to a limited extent. An initial test score would be more suited than the recommended track of secondary education. The researcher can only study what pupils learn at a school and what the added value of a school is if the entry level is determined in

this manner and multilevel data are available. An advantage of multilevel analysis is also that the distinction can be made between regression effects within a school and between schools for the benefit of parents. If parents seek to maximize the expected school achievements of their own child, they will want to order the schools according to a score which is corrected for the composition of the pupil population by means of within group, and not by means of between group regressions.

The present study also sheds light on the issue of inequality in the educational system. It appears that this issue requires a differentiated treatment. When we look into the *differences in achievement*, the large differences between pupils are striking. In year three of secondary education pupils of ethnic minorities perform less well than native-born pupils in Dutch and mathematics. In addition to ethnicity, there are also clear differences according to socioeconomic status. Adolescents from lower social classes have lower scores on the tests than pupils from higher classes. An inequality between the sexes is also apparent. Boys perform better in mathematics than girls. Girls perform better in Dutch than boys. To conclude, it may be stated that there are clear differences in these structural characteristics at the end of the school age.

When the *differences in progress* are studied, it appears that there is no influence of the characteristics of ethnicity and socioeconomic status. The only structural characteristic that has an effect on study progress is gender. The progress of girls in Dutch is larger than that of boys. In year three of secondary education significant differences in achievement become apparent, but these differences are hardly, if at all, caused by the pupils' progress in the first few years. Inequality according to ethnicity and socioeconomic status arises before pupils enter secondary education and does not increase between the first and the third year. In this period special policies to resolve the disadvantages for pupils with a different ethnic background or for pupils from lower social classes will, on average, be useless. The correction of disadvantages should focus mainly on the life style of the teenagers. No miracles should be expected of this type of policy. Cultural characteristics, such as the life style, do have an influence on differences in relative positions, but this influence is rather weak.

Geraadpleegde literatuur

- Agerbeek, M. (1998). Trouw-onderzoek schoolprestaties 1998. *Trouw*, 13 juni.
- Agerbeek, M. (1999a). *Keuzegids middelbare scholen 1999*. Zeist: Kerckebosch.
- Agerbeek, M. (1999b). Trouw-onderzoek schoolprestaties 1999. *Trouw*, 18 september.
- Agerbeek, M. (1999c). Wie geschoren wordt, moet stilzitten. *Trouw*, 11 mei, p.15.
- Agerbeek, M., Hageman, E., Kreulen, E. & Lakmaker, H. (1997). Trouw-onderzoek schoolprestaties. *Trouw*, 25 oktober.
- Ainsworth-Darnell, J.W. & Roscigno, V.J. (1997). *Race, cultural capital, and educational success: Persistent inequalities and disparate returns*. Toronto: Paper voor het congres van de American Sociological Association.
- Aitkin, M. & Longford, N. (1986). Statistical modelling in school effectiveness studies. With discussion *Journal of the Royal Statistical Society, A*, 149, 1-43.
- Aitkin, M. & Zuzovsky, R. (1994). Multilevel interaction models and their use in the analysis of large-scale school effectiveness studies. *School Effectiveness and School Improvement*, 5, 45-73.
- Allen, J., Heijke, H. & Velden, R. van der (1999). *Naar een nieuwe fase in het longitudinale onderwijsonderzoek*. Maastricht: ROA.
- Allison, P.D. (1990) Change scores as dependent variables in regression analysis. *Sociological Methodology*, 20, 93-114.
- Bandura, A. (1988). "Self-regulation of motivation and action through goal systems." In Hamilton, V., Bower, G.H. & Frijda, N.H. (eds.), *Cognitive perspectives on emotion and motivation* (pp. 37-61). Dordrecht: Kluwer.
- Bergh, H. van den & Kuhlemeier, H. (1997). Multi-level modellen voor de analyse van leerwinst vergeleken. *Tijdschrift voor Onderwijsresearch*, 22(2), 54-75.
- Berk, R.A. (1983) An introduction to sample selection bias in sociological data. *American Sociological Review*, 48(3), 386-398.
- Bernstein, B. (1971). *Class, codes and control. Theoretical studies towards a sociology of language*. London: Routledge.
- Bierman, K. (1996). "Family-school links: an overview." In Booth, A. & Dunn, J.F. (eds.). *Family-school links: how do they affect educational outcomes?* (pp. 275-287). Mahwah: Lawrence Erlbaum Associates.

- Blau, P.M. & Duncan, O.D. (1967). *The American occupational structure*. New York etc.: John Wiley & Sons.
- Block, J.H. & Burns, R.B. (1976). Mastery learning. *Review of Research in Education*, 4, 3-49.
- Bosker, R.J. (1990). *Extra kansen dankzij de school?* (Academisch proefschrift Rijksuniversiteit Groningen). Nijmegen: ITS.
- Bosker, R.J. (1991). De consistentie van schooleffecten in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 16(4), 206-218.
- Bosker, R.J. & Hofman, W.H.A. (1994). School effects on drop out. A multi-level logistic approach to assessing school-level correlates of drop out on ethnic minorities. *Tijdschrift voor Onderwijsresearch*, 19(1), 50-64.
- Bosker, R.J., Lam, J.F., Luyten, H., Steen, R. & Vos, H. de (1998). *Het vergelijken van scholen*. Enschede: Univesiteit Twente.
- Bosker, R.J. & Scheerens, J. (1989). Issues in the interpretation of the results of school effectiveness research. *International Journal of Educational Research*, 13(7), 741-751.
- Bosker, R.J. & Scheerens, J. (1999). Openbare prestatiegegevens; nuttigheid en validiteit. *Pedagogische Studiën*, 76(1), 61-73.
- Bosker, R.J., Velden, R.K.W. van der, & Hofman, W.H.A. (1985). *Een generatie geselecteerd. Deel 1: De loopbanen*. Groningen: RION.
- Bosman, R. (1993). *Opvoeden in je eentje: Een onderzoek naar de betekenis van het moedergezin voor de onderwijskansen van kinderen* (Academisch proefschrift Rijksuniversiteit Groningen). Lisse: Swets & Zeitlinger.
- Boudon, R. (1974). *Education, opportunity and social inequality: Changing prospects in western society*. New York etc.: John Wiley & Sons.
- Bourdieu, P. (1977). "Cultural reproduction and social reproduction". In Karabel, J. & Halsey, A.H. (eds.), *Power and ideology in education* (pp. 487-511). New York: Oxford University Press.
- Brandsma, H.P. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs* (Academisch proefschrift Rijksuniversiteit Groningen). Groningen: RION.
- Brandsma, H.P., Lugthart, E. & Werf, M.P.C. van der (1997). *Beschrijving van de school- en docentkenmerken in het eerste leerjaar van het voortgezet onderwijs: Een analyse op de eerste meting van VOCL'93*. Enschede/Groningen: OCTO/GION.
- Brandsma, H.P. & Werf, M.P.C. van der (1997). *Beschrijving van het onderwijspeil in het eerste leerjaar van het voortgezet onderwijs: Een analyse op de eerste meting van VOCL'93*. Enschede/Groningen: OCTO/GION.
- Brekelmans, M. (1989). *Interpersoonlijk gedrag van docenten in de klas*

- (Academisch proefschrift Universiteit Utrecht). Utrecht: WCC, 1989.
- Brink, W.P. van den & Koele, P. (1996³). *Statistiek. Deel 3: Toepassingen*. Meppel/Amsterdam: Boom.
- Brophy, J. & Good, T.L. (1986). "Teacher behavior and student achievement". In Wittrock, M.C. (ed.), *Handbook of research on teaching, third edition* (pp. 328-375). New York: MacMillan Inc.
- Bryk, A.S., Lee, V.E. & Holland, P.B. (1993). *Catholic schools and the common good*. Cambridge, MA: Harvard University Press.
- Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Bryk, A.S., Raudenbush, S.W., & Congdon, R.T. (1993). *HLM2 and HLM3. Computer programs and users' guide*. Chicago: University of Chicago.
- Carroll, J.B. (1962). "The prediction of success in intensive language training." In Glaser, R. (ed.), *Training research and education* (pp. 87-136). Pittsburgh: University of Pittsburgh Press.
- Carroll, J.B. (1963). A model of school learning. *Teachers College Record*, 64, 723-733.
- Carroll, J.B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher*, 18(1), 26-31.
- CBS (1995). *Zakboek onderwijsstatistieken 1995*. 's-Gravenhage: SDU.
- CBS (1997). *Zakboek onderwijsstatistieken 1997*. 's-Gravenhage: SDU.
- CBS (1998). *Jaarboek onderwijs 1998*. Alphen aan den Rijn: Samsom H.D. Tjeenk Willink.
- Chubb, J.E. & Moe, T.M. (1990). *Politics, markets, and America's schools*. Washington DC: Brookings Institution.
- Coleman, J.S. (1990). *Foundations of social theory*. Cambridge, MA/London: Belknap Press of Harvard University Press.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D. & York, R.L. (1966). *Equality of educational opportunity*. Washington DC: US Government Printing Office.
- Coleman, J.S. & Hoffer, T. (1987). *Public and private high schools: The impact of communities*. New York: Basic Books.
- Coleman, J.S., Hoffer, T. & Kilgore, S. (1981). *High school achievement: Public, Catholic, and private schools compared*. New York: Basic Books.
- Collins, L.M. (1996). Is reliability obsolete? A commentary on "Are simple gain scores obsolete?". *Applied Psychological Measurement*, 20(1), 289-292.
- Creemers, B.P.M. (1991). *Effectieve instructie: Een empirische bijdrage aan de*

- verbetering van het onderwijs in de klas*. 's-Gravenhage: SVO.
- Creemers, B.P.M. (1994). *The effective classroom*. London/New York: Cassell.
- Crul, M. (1996). Succesfactoren in de schoolloopbanen van Turkse en Marokkaanse jongeren van de Tweede generatie. *Tijdschrift voor Sociologie*, 17(3), 403-420.
- Dar, Y. & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal*, 23(3), 357-374.
- Dekkers, H.P.J.M. (1996). "Onderwijs en sekse". In Meijnen, G.W. (red.), *Onderwijsongelijkheid: Onderwijskundig Lexicon. Deel Centrale Onderwijsthema's* (pp. 45-60). Alphen aan den Rijn: Samsom H.D. Tjeenk Willink.
- Dekkers, H.P.J.M., Bosker, R.J. & Meijnen, G.W. (1998). *Het meritocratisch gehalte van het voortgezet onderwijs*. Programmabeschrijving van aanvraag voor aandachtsgebied bij Programmaraad Onderwijsonderzoek, NWO.
- Dekkers, H. & Wit, W. de (1995). Zeer voortijdig schoolverlaten in het voortgezet onderwijs. *Migrantenstudies*, 11(3), 203-218.
- Dijkstra, A.B. (1992). *De religieuze factor. Onderwijskansen en godsdienst: een vergelijkend onderzoek naar gereformeerd-vrijgemaakte scholen* (Academisch proefschrift Rijksuniversiteit Groningen). Nijmegen: ITS.
- Dijkstra, A.B. (1994). "Bijzonder en beter? Onderwijskansen in een verzuild bestel". In Miedema, S. & Klifman, H. (red.), *Christelijk onderwijs in ontwikkeling. Jaarboek 1994* (pp. 108-120). Kampen: Kok.
- Dijkstra, A.B. (1997). "Onderwijskansen en richting van de school". In Dijkstra, A.B., Dronkers, J. & Hofman, R. (red.), *Verzuiling in het onderwijs: Actuele verklaringen en analyse* (pp. 144-184). Groningen: Wolters-Noordhoff.
- Dijkstra, A.B. & Veenstra, R. (1999). *Functionele gemeenschappen, religiositeit en prestaties in het voortgezet onderwijs*. Te verschijnen.
- Dijkstra, A.B., Veenstra, R., Snijders, T., & Peschar, J. (1999). *Trouw moet blijken. Opmerkingen bij een hitlijst voor scholen*. Te verschijnen.
- Doyle, W. (1985). "Effective secondary classroom practices". In Kyle, R.M.J. (ed.), *Reaching for excellence: An effective schools sourcebook* (pp. 55-70). Washington: National Institute of Education.
- Downey, D.B. (1994). The school performance of children from single-mother and single-father families: Economic or interpersonal deprivation? *Journal of Family Issues*, 15(1), 129-147.
- Downey, D.B. (1995). Bigger is not better: Family size, parental resources, and children's educational performance. *American Sociological Review*, 60, 746-761.
- Drayer, E. (1997). Onrust in onderwijsland: De lumpsum in de praktijk. *Vrij*

- Nederland, 28 juni 1997, 23.
- Driessen, G. (1995). Het relatieve belang van sociaal milieu en etnische herkomst voor de verklaring van onderwijsachterstanden. *Tijdschrift voor Onderwijsresearch*, 20(4), 341-362.
- Driessen, G. & Werf, G. van der (1992). *Het functioneren van het voortgezet onderwijs: Beschrijving steekproef en psychometrische kwaliteit instrumenten*. Groningen/Nijmegen: RION/ITS.
- Dronkers, J. (1986). "Onderwijs en sociale ongelijkheid". In Kemenade, J.A. van, Lagerweij, N.A.J., Leune, J.M.G. & Ritzen, J.M.M. (red.), *Onderwijs - Bestel en beleid 2: onderwijs en samenleving A* (pp. 39-147). Groningen: Wolters-Noordhoff.
- Dronkers, J. (1990). De ontwikkelingen in het schoolloopbaanonderzoek: een terugblik op een decennium. *Tijdschrift voor Onderwijsresearch*, 15, 8-22.
- Dronkers, J. (1997a). Loopbanen door het voortgezet onderwijs in de jaren negentig. *Beleid en Maatschappij*, 22(2), 85-95.
- Dronkers, J. (1997b). Schoolcijfer zegt niet dat scholen met 5 onvoldoende functioneren. *Trouw*, 8 november, p. 12.
- Dronkers, J. (1997c). Zoals de ouden zongen, piepen de jongen. Intergenerationele overdracht van de kans op scheiding in Nederland. *Mens en Maatschappij*, 72(2), 149-165.
- Dronkers, J. (1998a). Discussie. Het betere is de vijand van het goede. Een reactie op de commentaren over het *Trouw* rapportcijfer. *Pedagogische Studiën*, 75(2), 142-150.
- Dronkers, J. (1998b). *The importance of cognitive abilities at primary school for educational and occupational success in the life course of a Dutch generation born around 1940*. Montreal: Paper voor het congres van de International Sociological Association.
- Dronkers, J. (1999). Het betere blijft nog steeds de vijand van het goede. Een antwoord op Van der Werf en Creemers, en Bosker en Scheerens. *Pedagogische Studiën*, 76(1), 74-82.
- Dronkers, J., Erp, M. van, Robijns, M. & Roeleveld, J. (1998). Krijgen leerlingen in de grote steden en met name in Amsterdam te hoge adviezen? *Tijdschrift voor Onderwijsresearch*, 23(1), 17-30.
- Duncan, O.D., Featherman, D.L. & Duncan, B. (1972). *Socio-economic background and achievement*. New York: Academic.
- Duncan, C., Jones, K. & Moon, G. (1996). "Blood pressure, age and gender". In Woodhouse, G. (ed.), *Multilevel modelling applications. A guide for users of MLn* (pp. 59-84). London: Institute of Education, University of London.

- Edmonds, R. (1979). Effective schools for the urban poor. *Educational leadership*, 37(1), 15-24.
- Eeden, P. van den, Jong, U. de, Koopman, P. & Roeleveld, J. (1993). *Schoolloopbanen in Amsterdam: Leerlingen op Amsterdamse scholen voor algemeen voortgezet onderwijs in de jaren zeventig*. Amsterdam/Lisse: Swets & Zeitlinger.
- Eijck, K. van (1999). *De invloed van persoonlijkheidskenmerken op de onderwijsloopbaan*. Utrecht: Paper voor de NSV Marktdag.
- Entwisle, D.R., Alexander, K.L. & Steffel Olson, L. (1997). *Children, schools, and inequalities*. Boulder: Westview Press.
- Fägerlind, I. (1975). *Formal education and adult earnings : a longitudinal study on the economic benefits of education*. Stockholm: Almqvist & Wiksell.
- Falbo, T. & Polit, D.F. (1986). Quantitative review of the only child literature: Research evidence and theory development. *Psychological Bulletin*, 100(2), 176-189.
- Fennessey, J. & Salganik, L.H. (1983). Credible comparison of instructional impact. The RAGS procedure. *Educational Measurement: Issues and Practice*, 11, 13-17.
- Fischer, C.S., Hout, M., Sánchez Jankowski, M., Lucas, S.R., Swidler, A. & Voss, K. (1996). *Inequality by design: Cracking the bell curve*. Princeton: Princeton University Press.
- Fraser, B.J., Walberg, H.J., Welch, W.W. & Hattie, J.A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11(2), 145-252.
- Freedman, D.A. (1991). Statistical models and shoe leather. *Sociological Methodology*, 21, 291-313.
- Gambetta, D. (1987). *Were they pushed or did they jump? Individual decision mechanisms in education*. Cambridge etc.: Cambridge University Press.
- Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review*, 57, 812-828.
- Ganzeboom, H.B.G. (1984). *Cultuur en informatieverwerking: Een empirisch-theoretisch onderzoek naar cultuurdeelname en esthetische waardering van architectuur* (Academisch proefschrift Universiteit Utrecht). Utrecht: s.n..
- Goldstein, H. (1995²). *Multilevel models in educational and social research*. London: Griffin.
- Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement*, 8(4), 369-395.

- Goldstein, H. & Sammons, P. (1997). The influence of secondary and junior schools on sixteen year examination performance: a cross-classified multilevel analysis. *School Effectiveness and School Improvement*, 8(2), 219-230.
- Goldstein, H., & Spiegelhalter, D.J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. With discussion. *Journal of the Royal Statistical Society, A*, 159, 385-443.
- Goor, A. van, Ploeg, J. van der, Tjeerdema, S. & Vries, M. de (1999). *Invloed van leerling- en gezinskenmerken op algemene vaardigheden in de basisvorming*. Groningen: s.n. (project onderwijs Vakgroep Sociologie, Rijksuniversiteit Groningen).
- Graaf, P.M. de (1987). *De invloed van financiële en culturele hulpbronnen in onderwijsloopbanen* (Academisch proefschrift Universiteit Utrecht). Nijmegen: ITS.
- Gray, J., Jesson, D., Goldstein, H., Hedger, K. & Rasbash, J. (1995). A multi-level analysis of school improvement: Changes in schools' performance over time. *School Effectiveness and School Improvement*, 6(2), 97-114.
- Grisay, A. (1996). *Evolution des acquis cognitifs et socio-affectifs des élèves au cours des années de college*. Liège: Université de Liège.
- Hacquebord, H. (1989). *Tekstbegrip van Turkse en Nederlandse leerlingen in het voortgezet onderwijs* (Academisch proefschrift Rijksuniversiteit Groningen). S.l.: s.n..
- Harris, J.R. (1998). *The nurture assumption : why children turn out the way they do*. New York etc.: Free Press.
- Heckmann, J.J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161.
- Heek, F. van *et al.* (1968) [1976³]. *Het verborgen talent: Milieu, schoolkeuze en schoolgeschiktheid*. Meppel: Boom.
- Hermans, H.J.M. (1983). *PMT-K-83. Prestatie Motivatie Test voor Kinderen. Handleiding*. Lisse: Swets & Zeitlinger.
- Hermans, J.J. (1981). *Niet-voortgezet onderwijs: Voortijdig schoolverlaten in het algemeen voortgezet onderwijs - omvang, aard en voorspelbaarheid*. Lisse: Swets & Zeitlinger.
- Herrnstein, R.J. & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: The Free Press.
- Heyl, E. (1996). *Het docentennetwerk. Structuur en invloed van collegiale contacten binnen scholen*. (Academisch proefschrift Universiteit Twente). S.l.: s.n..
- Heyns, B. (1978). *Summer learning and the effects of schooling*. New York etc.: Academic.

- Ho, E.S. & Willms, J.D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of Education*, 69(2), 126-141.
- Hoffer, T.B. & Coleman, J.S. (1990), "Changing families and communities: implications for schools". In: Mitchell, B. & Cunningham, L.L. (eds.) (1990), *Educational leadership and changing contexts of families, communities, and schools*. Chicago: University of Chicago Press, 118-134;
- Hoffer, T., Greeley, A.M. & Coleman, J.S. (1985). Achievement growth in public and Catholic schools. *Sociology of Education*, 58(2), 74-97.
- Hofman, R.H. (1993). *Effectief schoolbestuur: Een studie naar de bijdrage van schoolbesturen aan de effectiviteit van basisscholen* (Academisch proefschrift Rijksuniversiteit Groningen). Groningen: RION.
- Hoogendijk, W. & Wolfgram, P. (1995). *KEA Halverwege*. Rotterdam: CED.
- Horn, W. (1969). *Prüfsystem für Schul- und Bildungsberatung PSB*. Göttingen: Verlag für Psychologie.
- Hox, J.J. (1994). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.
- Huisman, M. (1999). *Item-nonresponse: occurrence, causes, and imputation of missing answers to test items*. Leiden: DSWO Press (dissertatie Rijksuniversiteit Groningen).
- Husén, T. with Emanuelsson, I., Fägerlind, I. & Liljefors, R. (1969). *Talent, opportunity and career. A twenty-six year follow-up of 1500 individuals*. Stockholm : Almqvist & Wiksell.
- Hustinx, P.W.J. (1996). De voorsprong van meisjes aan het begin van de schoolloopbaan in het voortgezet onderwijs. *Sociologische Gids*, 43(4), 264-279.
- Jager, B. de & Reezigt, G.J. (1996). *Onderwijseffectiviteit en metacognitieve vaardigheden*. Groningen: GION.
- Jager, P.K. (1998). Kwaliteitskaart verdoezelt waarheid. *Het Onderwijsblad*, 27 juni 1998, p. 14.
- Jencks, C., Smith, M., Ackland, H., Bane, M.J., Cohen, D., Gintis, H., Heyns, B. & Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York/London: Basic Books.
- Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39(1), 1-123.
- Jong, M.J. de (1989). "Ethnic culture, family background and educational attainment". In Bakker, B.F.M., Dronkers, J. & Meijnen, G.W. (eds.), *Educational opportunities in the welfare state: Longitudinal studies in educational and occupational attainment in The Netherlands* (pp. 59-72). Nijmegen: ITS.

- Jong, R. de & Westerhof, K.J. (1998). Huiswerk en leerprestaties. *Pedagogische Studiën*, 75(4), 262-276.
- Jungbluth, P., Peetsma, T. & Roeleveld, J. (1996). *Leerlingprestaties en leerlinggedrag in het primair onderwijs*. Ubbergen: Tandem Felix.
- Keith, T.Z. (1982). Time spent on homework and high school grades: a large-sample path analysis. *Journal of Educational Psychology*, 74(2), 248-253.
- Kenny, D.A. (1975). A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. *Psychological Bulletin*, 82(3), 345-362.
- Knuver, J.W.M. (1993). *De relatie tussen klas- en schoolkenmerken en het affectief functioneren van leerlingen* (Academisch proefschrift Rijksuniversiteit Groningen). Groningen: RION.
- Kohn, M.L. (1977²). *Class and conformity. A study in values. With a reassessment*. Chicago/London: University of Chicago Press.
- Kuiper, C.H. & Schuurman, M.I.M. (1981). *Scholieren van nu: Resultaten van onderzoek naar problemen van leerlingen, uitgevoerd op dertien scholen voor voortgezet onderwijs en lager beroepsonderwijs*. Leiden: Nederlands Instituut voor Praeventieve Gezondheidszorg/TNO.
- Kuyper, H. & Guldemon, H. (1997). *Studievaardigheden en huiswerkgedrag in de bovenbouw van havo en vwo*. Groningen: GION.
- Kuyper, H. & Swint, F.E. (1996). *Microscopisch schoolloopbaanonderzoek: De eerste drie jaren in het voortgezet onderwijs*. Groningen: GION.
- Kuyper, H. & Veenstra, D.R. (1995). *De effectiviteit van NT2-onderwijs aan gevorderde tweede taalleerders in het voortgezet onderwijs*. Groningen: GION (ongepubliceerd manuscript).
- Lam, S.F. (1997). *How the family influences children's academic achievement*. Hamden: Garland.
- Ledoux, G. (1996). "Onderwijs en etnische herkomst". In Meijnen, G.W. (red.), *Onderwijsongelijkheid: Onderwijskundig Lexicon. Deel Centrale Onderwijsthema's* (pp. 61-84). Alphen aan den Rijn: Samsom H.D. Tjeenk Willink.
- Lee, V.E. & Smith, J.B. (1997). High school size: Which works best and for whom? *Educational evaluation and policy analysis*, 19(3) 205-228.
- Leseman, P.P.M. (1990). *Structurele en pedagogische determinanten van schoolloopbanen* (Academisch proefschrift Erasmus Universiteit Rotterdam). Rotterdam: Rotterdamse Schooladviesdienst /Project Onderwijs en Sociaal Milieu.
- Leseman, P.P.M., Sijtsling, F.F. & Vries, E.M. de (1992). *Functionele ongeletterdheid in het lbo en mavo: Het effect van verschillen tussen scholen en*

- docenten*. Rotterdam: RISBO.
- Liere, C. van (1990). *Lastige leerlingen: Een empirisch onderzoek naar sociale oorzaken van probleemgedrag op basisscholen* (Academisch proefschrift Rijksuniversiteit Groningen). Amsterdam: Thesis.
- Linden, F.J. van der & Dijkman, T.A. (1989). *Jong zijn en volwassen worden in Nederland: Een onderzoek naar het psychosociaal functioneren in alledaagse situaties van de Nederlandse jongeren tussen 12 en 21 jaar*. Nijmegen: Hoogveld Instituut.
- Lord, F.M. (1960). Large sample covariance analysis when the control variable is fallible. *Journal of the American Statistical Association*, 55, 307-321.
- Lord, F.M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, 68(5), 304-305.
- Lucas, S.R. (1997). *Context and college entry*. Toronto: Paper voor het congres van de American Sociological Association.
- Maas, C.J.M. (1992). *Probleemleerlingen in het basisonderwijs* (Academisch proefschrift Universiteit Utrecht). Amsterdam: Thesis.
- Madaus, G.F., Kellaghan, T., Rakow, E.A. & King, D.J. (1979). The sensitivity of measures of school effectiveness. *Harvard Educational Review*, 49(2), 207-230.
- Marwijk Kooy-von Baumhauer, L. van (1984). *Scholen verschillen: Een verkennend vergelijkend onderzoek naar het intern functioneren van vijftientig scholengemeenschappen vwo-havo-mavo* (Academisch proefschrift Erasmus Universiteit Rotterdam). Groningen: Wolters-Noordhoff.
- Matthijssen, M.A.J.M. & Sonnemans, G.J.M. (1959). *Schoolkeuze en schoolsucces bij vmo en ulo in Noord-Brabant*. Tilburg: Zwijssen.
- Maxwell, S.E. & Delaney, H.D. (1990). *Designing experiments and analyzing data. A model comparison perspective*. Belmont, CA: Wadsworth.
- Meijnen, G.W. (1977). *Maatschappelijk achtergronden van intellectuele ontwikkeling* (Academisch proefschrift Rijksuniversiteit Groningen). Groningen: Wolters-Noordhoff.
- Meijnen, G.W. e.a. (1991). *Schoolvoorbeelden: Effectief onderwijs aan kinderen uit achterstandsmilieus*. Meppel: Edu'Actief.
- Muller, C. & Kerbow, D. (1993). "Parent involvement in the home, school, and community". In Schneider, B. & Coleman, J.S. (eds.), *Parents, their children and schools* (pp. 13-46). Boulder etc.: Westview Press.
- Neisser, U., Boodoo, G., Bouchard jr, T.J., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J. & Urbina, S. (1996).

- Intelligence: Knowns and Unknowns. *American Psychologist*, 51(2), 77-101.
- Niehof, J. (1997). *Resources and social reproduction: The effects of cultural and material resources on educational and occupational careers in industrial nations at the end of the twentieth century* (Academisch proefschrift Katholieke Universiteit Nijmegen). Amsterdam: Thesis.
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven/London: Yale University Press.
- OECD (1992, 1993, 1995, 1996, 1998). *Education at a glance. OECD Indicators*. Paris: OECD.
- Onderwijsraad (1998). *Kwaliteitskaart voortgezet onderwijs*. 's-Gravenhage: Onderwijsraad.
- Pallas, A.M. (1993). Schooling in the course of human lives: The social context of education and the transition to adulthood in industrial society. *Review of Educational Research*, 63(4), 409-447.
- Paterson, L. (1992). "Social class in Scottish education". In Brown, S. & Riddell, S. (eds.), *Class, race and gender in schools: A new agenda for policy and practice in Scottish education* (pp. 8-17). Edinburgh: Scottish Council for Research in Education.
- Peschar, J.L. (1975) [1987³]. *Milieu - School - Beroep: Een achteraf-experiment over de periode 1958-1973 naar de invloed van het sociaal milieu op de school- en beroepsloopbaan* (Academisch proefschrift Rijksuniversiteit Groningen). Groningen: Konstapel.
- Peschar, J.L. (1994). Waarover spraken zij? Enkele opmerkingen naar aanleiding van onderzoek naar klasse- en groepssamenstelling. *Pedagogisch Tijdschrift*, 19, 175-183.
- Peschar, J.L. & Veenstra, R. (1999). "Het Nederlandse onderwijssysteem". In Vis, J. & Moldenhauer, G. (red.), *Nederland en Duitsland. Wat Nederlanders en Duitsers van elkaar zouden moeten weten*. Münster: Waxmann. Te verschijnen.
- Peschar, J. & Wesselingh, A. (1995). *Onderwijssociologie*. Groningen: Wolters-Noordhoff.
- Petegem, P. van (1997). *Scholen op zoek naar hun kwaliteit. Effectieve-scholenonderzoek als inspiratiebron voor de zelfevaluatie van scholen*. Universiteit Gent: s.n. (dissertatie Universiteit Gent).
- Phillips, M. (1997). What makes schools effective? A comparison of the relationships of communitarian climate and academic climate to mathematics achievement and attendance during middle school. *American Educational*

- Research Journal*, 34(4), 633-662.
- Powell, A.G., Farrar, E. & Cohen, D.K. (1985). *The shopping mall school: Winners and losers in the educational marketplace*. Boston: Houghton Mifflin.
- Prick, L. (1997). Vergelijking van scholen is krakkemikkige vertoning. *NRC Handelsblad*, 1 november, p. 7.
- Raudenbush, S. & Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Reichardt, C.S. (1979). "The statistical analysis of data from nonequivalent group designs". In Cook, T.D. & Campbell, D.T. (eds), *Quasi-experimentation. Design and analysis issues for field settings* (pp. 147-205). Boston: Houghton Mifflin Company.
- Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.
- Rosenshine, B. & Stevens, R. (1986). "Teaching functions". In Wittrock, M.C. (ed.), *Handbook of research on teaching, third edition* (pp. 376-391). New York: MacMillan Inc.
- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York etc.: Holt, Rinehart & Winston.
- Rutter, M., Maughan, B., Mortimore, P. & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Somerset: Open Books.
- Ryan, B.A. & Adams, G.R. (1995). "The family-school relationships model". In Ryan, B.A., Adams, G.R., Gullotta, T.P., Weissberg, R.P. & Hampton, R.L. (eds.), *The family-school connection: Theory, research, and practice* (pp. 3-28). Thousand Oaks etc.: Sage.
- Sammons, P. (1995). Gender, ethnic and socio-economic differences in attainment and progress: A longitudinal analysis of student achievement over 9 years. *British Educational Research Journal*, 21(4), 465-485.
- Sammons, P., Hillman, J. & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. London: International School Effectiveness and Improvement Centre, Institute of Education.
- Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten*. 's-Gravenhage: SVO.
- Scheerens, J. (1992). *Effective schooling: Research, theory and practice*. London/New York: Cassell.
- Scheerens, J. & Bosker, R.J. (1997). *The foundations of educational effectiveness*.

- Oxford etc.: Pergamon.
- Schuyt, K. (1996). Kleinscholigheid. *de Volkskrant*, 4 juli, p. 11.
- Scott, R. & Walberg, H.J. (1979). Schools alone are insufficient: A response to Edmonds. *Educational Leadership*, 37(1), 24-27.
- Shavelson, R.J., Hubner, J.J. & Stanton, G.C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46(3), 407-441.
- Simons, P.R.J. (1995). "Leerlingenkenmerken". In Lowyck, J. & Verloop, N. (red.), *Onderwijskunde: Een kennisbasis voor professionals* (pp. 14-42). Groningen: Wolters-Noordhoff.
- Sizer, T.R. (1992). *Horace's school: Redesigning the American high school*. Boston/New York: Houghton Mifflin.
- Slavenburg, J.H. (1986). *Onderwijsstimulering en gezinsactivering* (Academisch proefschrift Rijksuniversiteit Groningen). 's-Gravenhage: SVO.
- Smith, T.E. (1993). Growth in academic achievement and teaching younger siblings. *Social Psychology Quarterly*, 56(1), 77-85.
- Snijders, T.A.B. & Bosker, R.J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. Newbury Park, CA: Sage.
- Steinberg, L. with Brown, B.B. & Dornbusch, S.M. (1996). *Beyond the classroom: Why school reform has failed and what parents need to do*. New York etc.: Simon & Schuster.
- Steinberg, L., Elmen, J.D. & Mounts, N.S. (1989). Authoritative parenting, psychosocial maturity, and academic success among adolescents. *Child Development*, 60, 1424-1436.
- Steinberg, L., Lamborn, S.D., Dornbusch, S.M. & Darling, N. (1992). Impact of parenting practices on adolescent achievement: authoritative parenting, school involvement, and encouragement to succeed. *Child Development*, 63, 1266-1281.
- Stevens, J. (1992²). *Applied multivariate statistics for the social sciences*. Hillsdale: Lawrence Erlbaum Associates.
- Stoel, W.G.R. (1980). *De relatie tussen de grootte van scholen voor voortgezet onderwijs en het welbevinden van de leerlingen. Deel 1: resultaten van een empirisch onderzoek in de Nederlandse onderwijssituatie*. Haren: RION.
- Swint, F. (1994). *Verschilcores: definitie, eigenschappen, toepassing en alternatieven*. Groninge: GION (interne notitie).
- Tesser, P.T.M. (1986). *Sociale herkomst en schoolloopbanen in het voortgezet onderwijs* (Academisch proefschrift Erasmus Universiteit Rotterdam). Nijmegen: ITS.
- Trier, U.P. & Peschar, J.L. (1995). "Cross curricular competencies: rationale and strategy for developing a new indicator". In Phillips, G. (ed.), *Measuring what*

- students learn* (pp. 99-108). Paris: OECD/CERI.
- Tuin, A.C. van der (1993). *Schoolkenmerken en effectiviteit van scholengemeenschappen*. Groningen: s.n. (doctoraal-scriptie Vakgroep Sociologie, Rijksuniversiteit Groningen).
- Tuijnman, A. (1989). *Recurrent education, earnings, and well-being. A fifty-year longitudinal study of a cohort of Swedish men*. Stockholm: Almqvist & Wiksell.
- Veenman, S. (1992). Effectieve instructie volgens het directe instructiemodel. *Pedagogische Studiën*, 69(4), 242-269.
- Veenstra, D.R. (1997). *Wat is er met onze school aan de hand? Acht portretten van scholen voor voortgezet onderwijs*. Groningen: ICS/Vakgroep Sociologie.
- Veenstra, D.R. (1998). Het mentoraat als aangrijpingspunt? Over de vormgeving van de eerste-lijnleerlingbegeleiding in het voortgezet onderwijs. *Pedagogische Studiën*, 75(4), 277-283.
- Veenstra, D.R. & Dijkstra, A.B. (1998). *Within private schools. The effect of religiosity on scholastic achievement in secondary education*. Te verschijnen.
- Veenstra, D.R., Dijkstra, A.B., Peschar, J.L. & Snijders, T.A.B. (1998). Discussie: Scholen op rapport. Een reactie op het Trouw-onderzoek naar schoolprestaties. *Pedagogische Studiën*, 75(2), 121-134.
- Veenstra, R. & Kuyper, H. (1998). Effectieve leerlingen en gezinnen. De invloed van structurele en culturele leerling- en gezinskenmerken op prestaties in het voortgezet onderwijs. *Mens en Maatschappij*, 73(3), 353-373.
- Velden, L.F.J. van der (1996). *Context, visie, aanpak en effectiviteit: De bestrijding van achterstanden van Nederlandse leerlingen in het basisonderwijs* (Academisch proefschrift Rijksuniversiteit Groningen). Groningen: GION.
- Velden, R.K.W. van der (1991). *Sociale herkomst en schoolsucces: Het effect van culturele en sociale hulpbronnen op leerlingen* (Academisch proefschrift Universiteit van Amsterdam). Groningen: RION.
- Verhelst, N.D., Glas, C.A.W. & Verstralen, H.H.F.M. (1995). *One-parameter logistic model (OPLM)*. Arnhem: CITO.
- Vos, H. de (1986). Een model van het interactiesysteem van de schoolklas. *Mens en Maatschappij*, 61(2), 147-168.
- Vries, A.M. de (1992). *Hoe breder, hoe beter? De effecten van heterogeen groeperen in het voortgezet onderwijs in Nederland* (Academisch proefschrift Rijksuniversiteit Groningen). Groningen: RION.
- Walberg, H.J. (1984). Improving the productivity of America's schools. *Educational leadership*, 41(8), 19-27.

- Wang, M.C., Haertel, G.D. & Walberg, H.J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249-294.
- Waslander, S. (1999). *Koopmanschap & burgerschap. Marktwerving in het onderwijs* (Academisch proefschrift Rijksuniversiteit Groningen). Assen: Van Gorcum.
- Weide, M.G. (1995). *Effectief basisonderwijs voor allochtone leerlingen* (Academisch proefschrift Rijksuniversiteit Groningen). Groningen: GION.
- Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology*, 71(1), 3-25.
- Werf, M.P.C. van der & Kuyper, H. (1998). *Oudergegevens VOCL 93-3. Respons, gemeten concepten en beschrijving van het databestand*. Groningen: GION.
- Werf, M.P.C. van der, Kuyper, H. & Lubbers, M.J. (1999). *Relaties tussen achtergrond- en gezinskenmerken van leerlingen en opbrengsten van het voortgezet onderwijs*. Groningen: GION.
- Werf, M.P.C. van der, Lubbers, M.J. & Kuyper, H. (1999). *Onderwijsresultaten van VOCL '89 en VOCL '93 leerlingen*. Groningen: GION.
- Werf, M.P.C. van der, Tuin, A.C. van der & Driessen, G. (1995). *Het functioneren van het voortgezet onderwijs: Kenmerken van scholen en docenten in het derde leerjaar*. Groningen/Nijmegen: GION/ITS.
- Willet, J.B. (1988). Questions and answers in the measurement of change. *Review of Research in Education*, 15, 345-422.
- Williams, R.H. & Zimmerman, D.W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement*, 20(1), 59-69.
- Willms, J.D. (1985). Catholic-school effects on academic achievement: new evidence from the high school and beyond follow-up study. *Sociology of Education*, 58(2), 98-114.
- Wilson, B.L. & Corcoran, T.B. (1988). *Successful secondary schools: Visions of excellence in American public education*. London etc.: The Falmer Press.
- Witte, J.F. (1992). Private school versus public school achievement: are there findings that should affect the educational choice debate? *Economics of Education Review*, 11, 371-394.
- Woodhouse, G., Rasbash, J., Goldstein, H. & Yang, M. (1996). "Introduction to multilevel modelling". In Woodhouse, G. (ed.), *Multilevel modelling applications. A guide for users of MLn* (pp. 9-57). London: Institute of Education, University of London.
- Zigler, E. & Muenchow, S. (1992). *Head Start: The inside story of America's most successful educational experiment*. New York: Basic Books.